

39

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
26 July 2001 (26.07.2001)

PCT

(10) International Publication Number
WO 01/53531 A2

- (51) International Patent Classification⁷: **C12Q 1/68**, C12N 15/12, C07K 14/47, C12N 5/10, A01K 67/027 // A61P 19/12
- (74) Agents: **FORBES, James, C. et al.**; Pharmacia Corporation, Corporate Patent Department, P.O. Box 5110, Chicago, IL 60680 (US).
- (21) International Application Number: **PCT/US01/00016**
- (22) International Filing Date: **18 January 2001 (18.01.2001)**
- (25) Filing Language: **English**
- (26) Publication Language: **English**
- (30) Priority Data:
60/176,523 18 January 2000 (18.01.2000) US
- (71) Applicant (for all designated States except US): **PHARMACIA CORPORATION** [US/US]; Corporate Patent Department, P.O. Box 5110, Chicago, IL 60680 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **PHIPPARD, Deborah** [GB/US]; 16938 Lewis Spring Farms, Wildwood, MO 63005 (US). **VASANTHAKAMUR, Geetha** [US/US]; 1176 Claytonia Terrace, St. Louis, MO 63117 (US). **DOTSON, Stanton** [US/US]; 379 Palomino Hill Court, Chesterfield, MO 63005 (US). **MA, Xiao-Jun** [CN/US]; 4482 Calle Mar De Armonia, San Diego, CA 92130 (US).
- Published:**
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



WO 01/53531 A2

(54) Title: **OSTEOARTHRITIS TISSUE DERIVED NUCLEIC ACIDS, POLYPEPTIDES, VECTORS, AND CELLS**

(57) Abstract: This invention relates to nucleic acids derived from osteoarthritis (OA) tissue as well as compounds and compositions made using these nucleic acids. A variety of nucleic acid and polypeptide compounds and compositions are described and specifically disclosed. The nucleic acids or polypeptides may be contained within vectors or host cells and then used to produce agents such as nucleic acids, polypeptides, fragments of polypeptides, antibodies, and variants of each. These molecules can be used to diagnose or treat osteoarthritis, or to analyze the disease-modifying activity of OA drugs. Cells containing one or more nucleic acids or polypeptides of the invention can also be used as targets in high-throughput screening methods, particularly in screening for compounds designed to identify compositions affecting osteoarthritis.

OSTEOARTHRITIS TISSUE DERIVED NUCLEIC ACIDS, POLYPEPTIDES, VECTORS, AND CELLS

Field of the Invention

5 This invention relates to nucleic acids derived from osteoarthritis (OA) tissue as well as compounds and compositions made using these nucleic acids. A variety of nucleic acid and polypeptide compounds and compositions are described and specifically disclosed. The nucleic acids or polypeptides may be contained within vectors or host cells and then used to produce agents such as
10 nucleic acids, polypeptides, fragments of polypeptides, antibodies, and variants of each. These molecules can be used to diagnose or treat osteoarthritis, or to analyze the disease-modifying activity of OA drugs. Cells containing one or more nucleic acids or polypeptides of the invention can also be used as targets in high-throughput screening methods, particularly in screening for compounds designed
15 to identify compositions affecting osteoarthritis.

Background of the Invention

Osteoarthritis (OA) is a slowly progressing degenerative disease characterized by cartilage destruction that affects one or more joints. It is the most prevalent articular disease, and can severely impair mobility and lower extremity
20 function (Ling and Bathon, *J Am Geriatr Soc* 46:216-25 (1998)). The term osteoarthritis implies an inflammatory disease. Although inflammatory cells may be present, OA is considered to be an intrinsic disease of cartilage in which biochemical and metabolic alterations result in its breakdown. Diagnosis is typically based upon radiological examination as well as clinical observations
25 such as localized tenderness and bony or soft tissue swelling, joint function, severity of pain and the ease with which everyday functions such as climbing stairs can be performed. Characteristic radiographic findings include subchondral bone sclerosis, subchondral cysts and osteophytosis. Although joint space narrowing is considered to be a marker for articular cartilage thinning, in patients

with early OA, who do not have radiographic evidence of bony changes, joint space narrowing alone does not accurately indicate the status of the articular cartilage. Similarly osteophytosis alone, in the absence of other radiographic features of OA, may be due to aging rather than OA. The correlation between the pathological severity of OA and symptoms is poor. Many individuals with radiographic changes of advanced OA have no symptoms. The risk factors for pain and disability in subjects with OA is still poorly understood.

Estimating OA disease progression presents an even greater challenge than diagnosis. The current "gold standard" measure of disease progression is the change in joint space caused by articular cartilage loss observed using plain X-rays (Mazucca et al., *Osteoarthritis and Cartilage* 5:217-226 (1997)). Since changes are small (1-2 mm per year), a minimum of one year is required before sufficient changes have occurred to be detectable. An additional problem with this technique is the requirement that the joint must be in exactly the same position for each radiological examination. Thus, there is a need in the art for improved OA diagnostic and disease progression monitoring-methods.

Current osteoarthritis therapies treat the associated pain and inflammation, but do little, if anything, to modify the course of the disease. Drug trials of potential disease modifying osteoarthritis drugs (DMOADs) require a minimum of one year before drug efficacy can be assessed. Therefore there is a great need for biological markers and design assay systems that will predict disease modifying efficacy of Osteoarthritis drugs (DMOADS) within a significantly shorter time frame than the current year required for assessment of radiological changes. Also, disease markers that will detect early disease and allow early intervention with pharmacologic agents are much needed, but they are presently unavailable (Ling and Bathon, *J Am Geriatr Soc* 46:216-25 (1998)).

Summary of the Invention

It is an object of the present invention to provide biological markers and diagnostic assays, which satisfy the unmet diagnostic and monitoring needs in the art. Markers of OA disease and for novel disease modifying therapeutics are described in this invention. These nucleic acid sequences have been identified using detailed genomics and bioinformatics analyses combined with our understanding of the pathology of OA. This invention identifies genes involved in the biology of OA and representing targets from which diagnostic tools, such as monoclonal antibodies, can be generated to aid in monitoring treatment, definition and diagnosis of OA as well as new targets for therapeutic drug intervention.

Exemplary Uses of the Invention

The present invention provides nucleic acids and polypeptide sequences that can be used to generate diagnostic tools and to develop small molecule drugs. The sequences of the present invention can also be used for the direct generation of therapeutics. Nucleic acid sequences of the present invention also represent probes and primers which facilitate cloning of genes.

Useful for generating diagnostic reagents

Nucleic acid probes, proteins, and polypeptides made using the nucleic acids of the present invention or derived from the sequence information of these nucleic acids are disclosed.

- a) Sequences represent markers that correlate with OA disease. Where ESTs are identified as coding for previously identified genes, full length sequence can be cloned (appropriate purification tags added if desired), expressed in a bacterial, yeast, insect, mammalian or other cell, and the recombinant protein used to raise antibodies which will have utility in OA diagnosis and may aid in analyzing DMOAD efficacy.

- 5 b) Sequences representing known genes that are now shown to correlate with OA disease allow commercially produced antibodies to these genes, raised for unrelated purposes, to be evaluated as markers of osteoarthritis disease and their response to drug treatment studied.
- 10 c) The full length sequences of the ESTs identified can also be directly injected into animals to elicit an antibody response.
- 15 d) Antibodies generated by any of the methods described above can be used to generate ELISAs (Enzyme linked immunosorbent assays) for the detection of OA serum or synovial fluid markers (or other appropriate body tissues and fluids e.g. saliva) and their response to drug treatment. The same antibodies are also useful for detecting OA and DMOAD markers using Western blotting techniques and other antibody-based diagnostics.
- 20 e) As described in a) proteins can be generated to the genes represented by the identified ESTs. This process will also generate positive controls to test the sensitivity and efficacy of the ELISAs described in d).
- 25 f) Where full length sequences are known or obtained, biochemical assays for the protein product can be designed to generate an assay for OA.

- 5
- g) The sequences specify oligonucleotides which can be used in polymerase chain reaction (PCR) detection procedures to determine the level of gene expression which infers the cellular state or gene response.
- 10
- h) PCR primer and probe sets homologous to the identified ESTs can be designed for use in quantitative PCR technology such as the TaqMan technology used in the ABI7700, therefore genes identified here as regulated as a result of OA can be quantified in human patients.
- 15
- i) Sequences can be used to clone their orthologous genes (if present) in species other than Homo sapiens. This allows their expression to be studied in animal models of OA such as the anterior cruciate ligament dog model, the rabbit partial meniscectomy model and the guinea pig spontaneous model.
- 20
- j) The DNA clones of these sequences can be spotted onto an array or microarray by themselves or as a set or in combination with other genes to determine the level of gene expression which can be used to infer a cellular state or cell response.
- 25
- k) Sequences can be placed on a computer-readable medium, and can be used to identify another nucleic acid by comparing nucleotide sequence information using computerized means.

Useful as targets for small molecule drug development

The present invention also provides methods and assays to discover compounds and compositions related to OA.

- 5 a) Sequences represent genes which can be cloned, over-expressed in a bacterial, yeast, insect, mammalian or other cell, and the active protein used in high throughput screening for novel inhibitors. For example, a biologically-active compound or composition may be identified by examining the interaction between a protein, polypeptide, or fragment of the invention and a compound or composition, and comparing this to a similar interaction of the protein, polypeptide, or fragment of the invention with a control.
- 10 b) Compounds and compositions which are detectable through screens using sequences that represent genes.
- 15 c) Sequences can be used to clone their promoter regions, which in turn can be linked to a reporter gene, such as luciferase, and the resultant recombinant reporter constructs used to screen compounds which alter gene expression.
- 20 d) The sequences can be used to identify transcription factors which modulate their expression. These transcription factors can be cloned, over-expressed in a bacterial, yeast, insect, mammalian or other cell, and the active transcription factor used in high throughput screening for small molecule inhibitors of gene expression.
- 25 e) Sequences represent cellular markers correlated to a cellular state which individually or in combination can be measured in response to a compound for those compounds which either suppress or

-7-

activate the genes and thus alter the state of the cell in a desired manner.

Useful for the direct generation of therapeutics

- 5 a) The sequences represent genes which can be cloned, over-expressed in a bacterial, yeast, insect, mammalian or other cell, and the encoded protein used as a protein therapeutic.
- 10 b) The sequences represent genes which can be directly injected to elicit antibodies useful as therapeutics.
- c) The sequences represent genes which can be cloned, expressed in a bacterial, yeast, insect, mammalian or other cell, and the protein used to generate antibodies useful as therapeutics.
- 15 d) The sequences can be introduced into an animal to create transgenic animals
- e) The sequences can be used to generate antisense DNA molecules useful to suppress gene expression and provide a therapeutic
20 benefit.
- f) The sequences can be used to generate antisense oligonucleotides useful to suppress gene expression and provide a therapeutic
25 benefit.
- g) The sequences can be used to generate sense DNA or sense oligonucleotides which will act by co-suppression to provide a therapeutic benefit.

- h) The sequences or genes represented by the sequences are useful as gene therapy for activating or suppressing themselves, other genes or entire pathways of genes.

5

The sequences will facilitate cloning the complete gene

The present invention also includes the entire coding region, the promoter which controls gene transcription, the untranslated region which may control RNA stability and translation as well as the genomic clone containing exon and intron information.

10

- a) The sequences specify oligonucleotide templates that can be used to amplify the full length gene using PCR.

15

- b) The DNA clones can be labeled in a manner that they can be used to hybridize to a corresponding full length gene in order to detect and clone the full length gene.

20

- c) The sequences have utility in other procedures, not limited to the two previous examples, to clone the full length gene.

Detailed Description of the Invention

A. General Concepts and Definitions

These detailed descriptions are presented for illustrative purposes only and are not intended as a restriction on the scope of the invention. Rather, they are merely some of the embodiments that one skilled in the art would

25

understand from the entire contents of this disclosure. All parts are by weight and temperatures are in Degrees centigrade unless otherwise indicated.

Abbreviations and Definitions

The following is a list of abbreviations and the corresponding meanings as

5 used interchangeably herein:

IMDM = Iscove's modified Dulbecco's media

mg = milligram

ml or mL = milliliter

µg or ug= microgram

10 µl or ul = microliter

ODNs= oligonucleotides

PCR= polymerase chain reaction

RP-HPLC = reverse phase high performance liquid chromatography

15 The following is a list definitions of various terms used herein:

The term "**altered**" means that expression differs from the expression response of cells or tissues not exhibiting the phenotype.

The term "**amino acid(s)**" means all naturally occurring L-amino acids.

20 The term "**biologically active**" means activity with respect to either a structural or a catalytic attribute, which includes the capacity of a nucleic acid to hybridize to another nucleic acid molecule, or the ability of a protein to be bound by an antibody (or to compete with another molecule for such binding), among others. Catalytic attributes involve the capacity of the agent to mediate a chemical reaction or response.

25 The term "**chromosome walking**" means a process of extending a genetic map by successive hybridization steps.

The term "**cluster**" means that BLAST scores from pairwise sequence comparisons of the member clones are similar enough to be considered identical with experimental error.

The term "**complement**" means that one nucleic acid exhibits complete
5 complementarity with another nucleic acid.

The term "**complementarity**" means that two molecules can hybridize to one another with sufficient stability to permit them to remain annealed to one another under conventional high stringency conditions.

The term "**complete complementarity**" means that every nucleotide of one
10 molecule is complementary to a nucleotide of another molecule.

The term "**degenerate**" means that two nucleic acid molecules encode for the same amino acid sequences but comprise different nucleotide sequences.

The term "**exogenous genetic material**" means any genetic material, whether naturally occurring or otherwise, from any source that is capable of being inserted
15 into any organism.

The term "**expansion**" means the differentiation and proliferation of cells.

The term "**expressed sequence tags (ESTs)**" means randomly sequenced members of a cDNA or complementary DNA library.

The term "**expression response**" means the mutation affecting the level or pattern
20 of the expression encoded in part or whole by one or more nucleic acid molecules.

The term "**fragment**" means a nucleic acid molecule whose sequence is shorter than the target or identified nucleic acid molecule and having the identical, the substantial complement, or the substantial homologue of at least 7 contiguous nucleotides of the target or identified nucleic acid molecule.

25 The term "**fusion molecule**" means a protein-encoding molecule or fragment that upon expression, produces a fusion protein.

The term "**fusion protein**" means a protein or fragment thereof that comprises one or more additional peptide regions not derived from that protein. Such molecules

may be derivatized to contain carbohydrate or other moieties (such as keyhole limpet hemocyanin, etc.).

The term “**hybridization probe**” means any nucleic acid capable of being labeled and forming a double-stranded structure with another nucleic acid over a region
5 large enough for the double stranded structure to be detected.

The term “**isolated**” means an agent is separated from another specific component with which it occurred. For example, the isolate material may be purified to essential homogeneity, as determined by PAGE or column chromatography, such as HPLC. An isolated nucleic acid can comprise at least about 50, 80, or 90% (on
10 a molar basis) of all macromolecular species present. Some of these methods described later lead to degrees of purification appropriate to identify single bands in electrophoresis gels. However, this degree of purification is not required.

The term “**marker nucleic acid**” means a nucleic acid molecule that is utilized to determine an attribute or feature (*e.g.*, presence or absence, location, correlation,
15 etc.) of a molecule, cell, or tissue.

The term “**mimetic**” refers to a compound having similar functional and/or structural properties to another known compound or a particular fragment of that known compound.

The term “**minimum complementarity**” means that two molecules can hybridize
20 to one another with sufficient stability to permit them to remain annealed to one another under at least conventional low stringency conditions.

The term “**PCR probe**” means a nucleic acid capable of initiating a polymerase activity while in a double-stranded structure with another nucleic acid. For example, Krzesicki, *et al.*, *Am. J. Respir. Cell Mol. Biol.* 16:693-701 (1997),
25 incorporated by reference in its entirety, discusses the preparation of PCR probes for use in identifying nucleic acids of osteoarthritis tissue. Other methods for determining the structure of PCR probes and PCR techniques have been described.

The term "**phenotype**" means any of one or more characteristics of an organism, tissue, or cell.

The term "**polymorphism**" means a variation or difference in the sequence of the gene or its flanking regions that arises in some of the members of a species.

- 5 The term "**primer**" means a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis under appropriate conditions (e.g., in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a
10 primer depends on the intended use of the primer, but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template, but must be sufficiently complementary to hybridize with a template.

- 15 The term "**probe**" means an agent that is utilized to determine an attribute or feature (e.g. presence or absence, location, correlation, etc.) of a molecule, cell, tissue, or organism.

- The term "**product score**" refers to a formula which indicates the strength of a BLAST match using the fraction of overlap of two sequences and the percent
20 identity. The formula is as follows:

$$\text{Product Score} = \frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length}(\text{Seq1}), \text{length}(\text{Seq2})\}}$$

- The term "**promoter region**" means a region of a nucleic acid that is capable,
25 when located in *cis* to a nucleic acid sequence that encodes for a protein or peptide, of functioning in a way that directs expression of one or more mRNA molecules.

The term "**protein fragment**" means a peptide or polypeptide molecule whose amino acid sequence comprises a subset of the amino acid sequence of that protein.

The term "**protein molecule/peptide molecule**" means any molecule that
5 comprises five or more amino acids.

The term "**recombinant**" means any agent (e.g., DNA, peptide, etc.), that is, or results from, however indirectly, human manipulation of a nucleic acid molecule. The recombination may occur inside a cell or in a tube.

The term "**selectable or screenable marker genes**" means genes whose
10 expression can be detected by a probe as a means of identifying or selecting for transformed cells.

The term "**singleton**" means a single clone.

The term "**specifically bind**" means that the binding of an antibody or peptide is not competitively inhibited by the presence of non-related molecules.

15 The term "**specifically hybridizing**" means that two nucleic acid molecules are capable of forming an anti-parallel, double-stranded nucleic acid structure.

The term "**substantial complement**" means that a nucleic acid sequence shares at least 80% sequence identity with the complement.

The term "**substantial fragment**" means a fragment which comprises at least 100
20 nucleotides.

The term "**substantial homologue**" means that a nucleic acid molecule shares at least 80% sequence identity with another.

The term "**substantial identity**" means that 70% to about 99% of a region or fragment in a molecule is identical to a region of a different molecule. When the
25 individual units (e.g., nucleotides or amino acids) of the two molecules are schematically positioned to exhibit the highest number of units in the same position over a specific region, a percentage identity of the units identical over the total number of units in the region is determined. Numerous algorithmic and

computerized means for determining a percentage identity are known in the art. These means may allow for gaps in the region being considered in order to produce the highest percentage identity.

The term “**substantially hybridizing**” means that two nucleic acid molecules can
5 form an anti-parallel, double-stranded nucleic acid structure under conditions (e.g. salt and temperature) that permit hybridization of sequences that exhibit 90% sequence identity or greater with each other and exhibit this identity for at least a contiguous 50 nucleotides of the nucleic acid molecules.

The term “**substantially purified**” means that one or more molecules that are or
10 may be present in a naturally occurring preparation containing the target molecule will have been removed or reduced in concentration.

The term “**tissue sample**” means any sample that comprises more than one cell.

B. Data Mining and Bioinformatic Computer Analysis

The present invention relates to osteoarthritis tissue-derived nucleic acid
15 sequences SEQ NOs. 1-82. These targets were identified by bioinformatic computer analysis and data mining of expressed sequence tags (ESTs) derived from sequencing cDNAs from normal and OA diseased tissues. The data mining effort used sequence comparison techniques (based on BLAST computer program comparisons of individual ESTs) to evaluate which ESTs were preferentially
20 observed in the target libraries versus control and/or normal libraries. The selected set of sequences represent a group of genes encoding multiple molecular targets for OA diagnostics. This same set of sequences also represents targets for new OA therapeutic drugs.

The present invention also relates to sequences derived from SEQ NOs. 1-
25 82. A variety of computerized means for identifying sequences derived from the SEQ NOs. 1-82 exists. These include the five implementations of BLAST, three designed for nucleotide sequences queries (BLASTN, BLASTX, and TBLASTX) and two designed for protein sequence queries (BLASTP and TBLASTN), as well

as FASTA and others (Coulson, *Trends in Biotechnology* 12:76-80 (1994); Birren *et al.*, *Genome Analysis* 1:543-559 (1997)).

The BLASTX is one such program that allows the comparison of nucleic acid sequences in the present invention to protein databases to search for the presence of protein homologs. The strength of a BLAST identity match is indicated by the "product score". This score is a normalized value between 0 and 100, with 100 indicating 100% identity over the entire length of the shorter of the two sequences, and 0 representing no shared identity between the sequences.

Other programs which use either individual sequences or make models from related sequences to further identify sequences derived from SEQ NO 1- SEQ NO 82 exist. Model building and searching programs includes HMMer (Eddy), MEME (Bailey and Elkan, *Ismb* 3: 21-29 (1995)) and PSI-BLAST (Altschul *et al.*, *Nucleic Acids Res* 25: 3389-3402 (1997)). Another set of programs which use predicted, related, or known protein structures to further identify sequences derived from SEQ NO 1- SEQ NO 82 exists. Structure-based searching programs includes ORF and PROSITE (Bairoch, A. *Nucleic Acid Res.* 20(Suppl.):2013-18 (1992)). Other programs which use individual sequences or related groups of sequences rely on pattern discovery to further identify sequences derived from SEQ NO:1- 82 exist. Pattern recognition programs include Teiresias (Rigoutsos, I. and A. Floratos, *Bioinformatics* 14:55-67 (1998)), Genotator (Harris, N. *Genome Res.* 7:754-762 (1997)).

For identification of protein structure, SPASM, RIGOR, and many other programs exist (Kleywegt, G. *J. Mol. Biol.* 285:1887-97 (1999)). Guernier *et al.* *Bioinformatics* 15:413-21 (1999)). These programs can search any appropriate database, such as but not limited to GenBank, dbEST, EMBL, SwissProt, PIR, and GENES, or protein databases such Protein DataBank (PDB), SCOP, CATH, the Protein Information Resource (PIR) International Protein Sequence Database, the RESAID Database, and FSSP. Computerized means for designing modifications

in protein structure are also known in the art (Dahiyat and Mayo, *Science* 278:82-87 (1997)). Additional information about protein characteristics also can be predicted, such as detecting signal peptides which are secreted (SignalP program) (Nielsen et al., *Protein Engineering* 12:3-9 (1999)).

5 The following protein or polypeptide embodiments of the invention can be identified through assays known in the art, including high throughput screening assays. The proteins or polypeptides possess a detectable activity in a functional assay and can be identified by that functional assay. For example, a kinase activity assay is discussed in U.S. Patent 5,759,787, and the references therein.

10 **C. Agents of the Invention**

(a) Nucleic Acids

 Agents of the present invention include nucleic acids and, more specifically, osteoarthritis tissue-derived nucleic acids. A subset of the nucleic acid molecules of the invention includes nucleic acids that are associated with a
15 gene or fragment thereof. Another subset of the nucleic acids of the invention includes those that encode proteins, polypeptides, or fragments of proteins or polypeptides. In one embodiment, the invention includes nucleic acid sequences having SEQ NOs. 1-82. In one embodiment, the invention includes the sequences identified in Table 1. In another embodiment, the nucleic acids of the invention
20 are derived from one or more EST sequences identified in Table 2.

 Fragment nucleic acids may encompass significant portion(s) of, or indeed most of, these nucleic acids. For example, a fragment nucleic acid can encompass an OA marker gene homolog or fragment thereof. Alternatively, the fragments may comprise smaller oligonucleotides (for example, having from about 7 to
25 about 250 nucleotides, or from about 15 to about 30 nucleotide).

 Nucleic acids or fragments thereof of the invention are capable of specifically hybridizing to other nucleic acids under certain circumstances. In one embodiment, a nucleic acid of the present invention will specifically hybridize to

one or more of the nucleic acids set forth in SEQ NO: 1 through SEQ NO: 82, or complements thereof, under moderately stringent conditions, for example at about 2.0 X SSC and about 65°C.

Conventional stringency conditions are described by Sambrook, *et al.*,
5 *Molecular Cloning, A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Press, Cold Spring Harbor, New York (1989), and by Haymes, *et al. Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, DC (1985). Departures from complete complementarity are therefore permissible, as long as such departures do not completely preclude the capacity of the molecules to form
10 a double-stranded structure. Thus, in order for a nucleic acid molecule to serve as a primer or probe it need only be sufficiently complementary in sequence to be able to form a stable double-stranded structure under the particular solvent and salt concentrations employed.

Appropriate stringency conditions that promote DNA hybridization, for
15 example, 6.0 X sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 X SSC at 50°C, are known to those skilled in the art or can be found in Ausubel, *et al.*, *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989) (see especially sections 6.3.1-6.3.6). [This reference and the supplements through December 1999 can be relied to make or use any
20 embodiment of the invention.] For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 X SSC at 50°C to a high stringency of about 0.2 X SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C. Temperature and salt
25 conditions may be varied independently.

In a particularly preferred embodiment, a nucleic acid of the invention will include those nucleic acids that specifically hybridize to one or more of the nucleic acids set forth in SEQ NO: 1 through SEQ NO: 82, or complements

thereof, under high stringency conditions. In one aspect of the present invention, the nucleic acid molecules of the present invention comprise one or more of the nucleic acid sequences set forth in SEQ NO: 1 through to SEQ NO: 82, or complements thereof.

5 In another aspect of the invention, one or more of the nucleic acid molecules of the present invention share between 80% and 100% sequence identity, or alternatively from between 90% to 100% sequence identity, with one or more of the nucleic acid sequences set forth in SEQ NO: 1 through to SEQ NO: 82 or complements thereof.

10 **(i) Nucleic Acids Comprising Genes or Fragments
Thereof**

This invention also provides genes corresponding to the cDNA sequences disclosed herein, also called OA Marker nucleic acids. The corresponding genes can be isolated in accordance with known methods using the sequence information
15 disclosed herein. The methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials.

The invention provides naturally existing gene homologues or fragments thereof. Genomic sequences can be screened for the presence of protein
20 homologues utilizing one or a number of different search algorithms that have been developed, such as the suite of BLAST programs.

In one embodiment of the present invention, the homologue protein or fragment thereof exhibits a BLASTX probability score of less than 1E-30, or alternatively a BLASTX probability score of between about 1E-30 and about 1E-
25 12, or alternatively a BLASTX probability score of greater than 1E-12 with a nucleic acid or gene of this invention. In another embodiment of the present invention, the nucleic acid molecule encoding the gene homologue or fragment thereof exhibits a % identity with its homologue of between about 25% and about

99%. In a further embodiment, the gene homologue or fragment has a single nucleotide difference from its homologue.

In another preferred embodiment, nucleic acid molecules having SEQ NO: 1 through SEQ NO: 82, or complements and fragments of either, can be utilized to
5 obtain homologues equivalent to the naturally existing homologues. The degeneracy of the genetic code, which allows different nucleic acid sequences to code for the same protein or peptide, is known in the literature (see U.S. Patent No. 4,757,006). As used herein a nucleic acid molecule is degenerate of another
10 nucleic acid molecule when the nucleic acid molecules encode for the same amino acid sequences but comprise different nucleotide sequences. An aspect of the present invention is that the nucleic acid molecules of the present invention include nucleic acid molecules that are degenerate of those set forth in SEQ NO: 1 through to SEQ NO: 82 or complements thereof.

-20-

In a further aspect of the present invention, one or more of the nucleic acid molecules of the present invention differ in nucleic acid sequence from those encoding a homologue or fragment thereof in SEQ NO: 1 through SEQ NO: 82, or complements thereof, due to the degeneracy in the genetic code in that they

5 encode the same protein but differ in nucleic acid sequence. In another further aspect of the present invention, one or more of the nucleic acid molecules of the present invention differ in nucleic acid sequence from those encoding an homologue of fragment thereof in SEQ NO: 1 through SEQ NO: 82, or complements thereof, due to fact that the different nucleic acid sequence encodes a

10 protein having one or more conservative amino acid residue. Examples of conservative substitutions are set forth below. Codons capable of coding for such conservative

	<u>Original Residue</u>	<u>Conservative Substitutions</u>
	Ala	ser
15	Arg	lys
	Asn	gln; his
	Asp	glu
	Cys	ser; ala
	Gln	asn
20	Glu	asp
	Gly	pro
	His	asn; gln
	Ile	leu; val
	Leu	ile; val
25	Lys	arg; gln; glu
	Met	leu; ile
	Phe	met; leu; tyr
	Ser	thr

-21-

Thr	ser
Trp	tyr
Tyr	trp; phe
Val	ile; leu

5

(ii) **Nucleic Acids Comprising Regulatory Elements**

One class of agents of the invention includes nucleic acids having promoter regions or partial promoter regions or regulatory elements. Promoter regions are typically found upstream of the trinucleotide ATG sequence at the start site of a protein coding region. The nucleic acids of the invention may be used to isolate promoters of cell-enhanced, cell-specific, tissue-enhanced, tissue-specific, developmentally- or physiologically-regulated expression profiles. Isolation and functional analysis of the 5' flanking promoter sequences from genomic libraries, for example, using genomic screening methods and PCR techniques, results in the isolation of useful promoters and transcriptional regulatory elements. These methods are known to those of skill in the art and have been described (see, for example, Birren *et al.*, *Genome Analysis: Analyzing DNA*, 1, (1997), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.).

For example, in one embodiment, a regulatory element is detected by incubating nucleic acid(s), or preferably fragments such as ESTs, with members of genomic libraries (*e.g.*, synovial tissue libraries) and recovering clones that hybridize to the nucleic acid(s). Sequencing techniques can then identify regulatory elements from known sequence motifs or known assays for detecting regulatory sequences within a certain proximity to transcription and translation start and stop sites can be used. In a second embodiment, methods of "chromosome walking," or inverse PCR may be used to obtain regulatory elements (Frohman *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 85:8998-9002 (1988); Ohara *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 86: 5673-5677 (1989); Pang *et al.*,

Biotechniques 22(6): 1046-1048 (1977); Huang *et al.*, *Methods Mol. Biol.* 69: 89-96 (1997); Huang, *et al.*, *Method Mol. Biol.* 67:287-294 (1997); Benkel *et al.*, *Genet. Anal.* 13: 123-127 (1996); Hartl *et al.*, *Methods Mol. Biol.* 58: 293-301 (1996)).

5 Promoters and regulatory elements obtained utilizing the nucleic acids of the invention can also be modified to affect their control characteristics. Examples of these modifications include, but are not limited to, enhancer sequences as reported by Kay *et al.*, *Science* 236:1299 (1987). Genetic elements such as these can be used to enhance gene expression of new and existing proteins
10 or polypeptides.

(b) Proteins and Polypeptides

This invention also provides a compound or composition comprising one or more polypeptides, which comprise: 1) at least one fragment, segment, or domain of at least 15-1,000 contiguous amino acids, with at least one portion
15 encoded by one or more of SEQ NOS: 1-82; 2) at least one amino acid sequence selected from those encoded by at least one of SEQ NOS: 1-82; or 3) at least one modification corresponding to fragments, segments, or domains within one of SEQ NOS:1- 82.

Protein and peptide molecules can be identified using known protein or
20 peptide molecules as a target sequence or target motif in the BLAST programs of the present invention. Proteins or peptides may undergo a variety of modifications, including post-translational modifications, such as disulfide bond formation, glycosylation, phosphorylation, or oligomerization. The term "protein" or "polypeptide" includes any protein molecule that is modified by any biological
25 or non-biological process. Proteins encoded by sequences of the present invention can also be fusion proteins. A fusion protein or peptide molecule of the present invention is preferably produced via recombinant means.

Another class of agents comprises protein or peptide molecules encoded by SEQ NO: 1 through SEQ NO: 82 or complements thereof or, fragments or fusions thereof in which conservative, non-essential, or not relevant, amino acid residues have been added, replaced, or deleted. An example is the homologue protein of an
5 OA marker protein. Such a homologue can be obtained by any of a variety of methods. For example, as indicated above, one or more of the disclosed sequences (SEQ NO: 1 through SEQ NO: 82, or complements thereof) will be used to define a pair of primers that may be used to isolate the homologue-encoding nucleic acid molecules from any desired species. Such molecules can be
10 expressed to yield homologs by recombinant means.

Proteins or polypeptides of the invention can be expressed as variants that facilitate purification. For example, a fusion protein to such proteins as maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX) are known in the art [New England BioLab, Beverly, Mass., Pharmacia, Piscataway,
15 N.J., and InVitrogen, San Diego, CA]. The polypeptide or protein can also be a tagged variant to facilitate purification, such as with histidine or methionine rich regions [His-Tag; available from LifeTechnologies Inc, Gaithersburg, MD] that bind to metal ion affinity chromatography columns, or with an epitope that binds to a specific antibody [Flag, available from Kodak, New Haven, Conn.]. An
20 exemplary, non-limiting list of commercially available vectors suitable for fusion protein expression includes: pBR322 (Promega); pGEX (Amersham); pT7 (USB); pET (Novagen); pIBI (IBI); pProEX-1 (Gibco/BRL); pBluescript II (Stratagene); pTZ18R and pTZ19R (USB); pSE420 (Invitrogen); pAc360 (Invitrogen); pBlueBac (Invitrogen); pBacPAK (Clontech); pHIL (Invitrogen); pYES2
25 (Invitrogen); pCDNA (Invitrogen); and pREP (Invitrogen).

A number of other purification methods or means are also known and can be used. These include, but are not limited to, anion or cation exchange chromatography, ethanol precipitation, affinity chromatography, high performance

liquid chromatography (HPLC), and reverse-phase high performance liquid chromatography (RP-HPLC), optionally employing hydrophobic RP-HPLC media, e.g., silica gel, to further purify the protein. Combinations of methods and means can also be employed to provide a substantially purified recombinant polypeptide or protein. The particular method used will depend upon the properties of the polypeptide, and the selection of the host cell; appropriate methods will be readily apparent to those skilled in the art.

The polypeptide or protein of the invention may also be expressed via transgenic animals. Methods and means employing the milk of transgenic domestic animals are known in the art.

One or more of the proteins, polypeptides, or fragments may be produced via chemical synthesis. Methods for synthetic construction are known to those skilled in the art. The synthetically-constructed sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with proteins, may possess biological properties in common, including protein activity. Thus, they may be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

(c) Antibodies

One aspect of the present invention concerns antibodies, single-chain antigen binding molecules, or other proteins that specifically bind to one or more of the protein or peptide molecules of the present invention and their homologues, fusions or fragments. Such antibodies may be used to quantitatively or qualitatively detect the protein or peptide molecules of the present invention. As used herein, an antibody or peptide is said to "specifically bind" to a protein or peptide molecule of the present invention if such binding is not competitively inhibited by the presence of non-related molecules.

Nucleic acid molecules that encode all or part of the protein of the present invention can be expressed, by recombinant means, to yield protein or peptides that can in turn be used to elicit antibodies that are capable of binding the expressed protein or peptide. Such antibodies may be used in immunoassays for that protein. Such protein-encoding molecules or their fragments may be a "fusion" molecule (*i.e.*, a part of a larger nucleic acid molecule) such that, upon expression, a fusion protein is produced. It is understood that any of the nucleic acid molecules of the present invention may be expressed, by recombinant means, to yield proteins or peptides encoded by these nucleic acid molecules.

1.0' The antibodies that specifically bind proteins and protein fragments of the present invention may be polyclonal or monoclonal, and may comprise intact immunoglobulins, or antigen binding portions of immunoglobulins (such as (F(ab'), F(ab')₂ fragments), or single-chain immunoglobulins producible, for example, via recombinant means. Conditions and procedures for the construction, manipulation and isolation of antibodies (see, for example, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, New York (1988)) are well known in the art.

Additionally, methods of raising antibodies to polypeptides of the present invention are well known in the art (Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, New York (1988)). Animals are immunized with purified protein (or fragment thereof) in appropriate amounts and using standard schedules, and are bled and the blood screened for the presence of anti-protein or peptide antibodies using, for example, a direct binding Enzyme-Linked Immunoassay (ELISA). In one embodiment, anti-protein or peptide monoclonal antibodies are isolated using a fusion of a protein, protein fragment, or peptide of the present invention, or conjugate of a protein, protein fragment, or peptide of the present invention, as immunogens. It is understood

that any of the protein or peptide molecules of the present invention may be used to raise antibodies.

As discussed below, such antibody molecules or their fragments may be used for diagnostic purposes. Where the antibodies are intended for diagnostic
5 purposes, it may be desirable to derivatize them, for example with a ligand group (such as biotin) or a detectable marker group (such as a fluorescent group, a radioisotope or an enzyme).

The ability to produce antibodies that bind the protein or peptide molecules of the present invention permits the identification of mimetic compounds of those
10 molecules. Mimetic compounds can be synthesized chemically. Combinatorial chemistry techniques, for example, can be used to produce libraries of peptides (see WO 9700267), polyketides (see WO 960968), peptide analogues (see WO 9635781, WO 9635122, and WO 9640732), oligonucleotides for use as mimetic compounds derived from this invention. Mimetic compounds and libraries can
15 also be generated through recombinant DNA-derived techniques. For example, phage display libraries (see WO 9709436), DNA shuffling (see US Patent 5,811,238) other directed or random mutagenesis techniques can produce libraries of expressed mimetic compounds. It is understood that any of the agents of the present invention can be substantially purified and/or be biologically active and/or
20 recombinant.

(d) Transformed Cells

A recombinant protein may be produced by operably linking a regulatory control sequence to a nucleic acid of the present invention and putting it into an expression vector. Regulatory sequences include promoters, enhancers, and other
25 expression control elements which are described in Goeddel (*Gene Expression Technology: Methods in Enzymology* 185. Academic Press, San Diego, CA (1990)). For example, the native regulatory sequences or regulatory sequences native to the transformed host cell can be used. One of skill in the art is familiar

with numerous examples of these additional functional sequences, as well as other functional sequences, that may optionally be included in an expression vector. The design of the expression vector may depend on such factors as the choice of the host cell to be transformed, and/or the type of protein desired. Many such
5 vectors are commercially available, including linear or enclosed elements (see for example, Broach, et al., *Experimental Manipulation of Gene Expression*, ed. M. Inouye, Academic Press, (1983); Sambrook, et al., *Molecular Cloning, A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Press, Cold Spring Harbor, New York (1989)). Typically, expression constructs will contain one or more
10 selectable markers, including the gene that encodes dihydrofolate reductase and the genes that confer resistance to neomycin, tetracycline, ampicillin, chloramphenicol, kanamycin and streptomycin resistance.

Prokaryotic and eukaryotic host cells transfected by the described vectors are also provided by this invention. For instance, cells which can be transfected
15 with the vectors of the present invention include, but are not limited to, bacterial cells such as *E. coli* (e.g., *E. coli* K 12 strains), *Streptomyces*, *Pseudomonas*, *Serratia marcescens* and *Salmonella typhimurium*, insect cells (baculovirus), including *Drosophila*, fungal cells, such as yeast cells, plant cells, and ovary cells (CHO), and COS cells.

20 Thus, the nucleic acid molecules described can be used to produce a recombinant form of the protein via microbial or eukaryotic cellular processes. Ligating the polynucleic acid molecule into a gene construct, such as an expression vector, and transforming or transfecting into hosts, either eukaryotic (yeast, avian, insect, plant, or mammalian) or prokaryotic (bacterial cells), are
25 standard procedures used in producing other well known proteins. Similar procedures, or modifications thereof, can be employed to prepare recombinant proteins according to the present invention by microbial means or tissue-culture

technology. Accordingly, the invention pertains to the production of encoded proteins or polypeptides by recombinant technologies.

D. Uses of the Agents of the Invention

1. Methods for Identifying Bioactive Proteins, Polypeptides, Fragments, or Variants of the Invention

5

Once the nucleic acid has been used to produce a protein, polypeptide, or a variant or fragment, any one of a number of assays can be used to identify bioactivity.

In addition, the agents of the invention are especially useful in high throughput screening methods. In general, these methods involve individual sample assay volumes less than about 250 μ l, or more preferably less than about 100 μ l. With smaller sample volumes, numerous individual assays can be performed simultaneously and via computer-operated instrumentation. The assays comprise the detectable interaction between a protein, polypeptide, fragment, nucleic acid, or antibody of the invention (sometimes referred to as the target) and an assay compound. Thus, the assays comprise two components, a target and an assay compound, where the assay compound may be part of a composition of multiple compounds. It is also possible for the agents of this invention to be used as assay compounds in screening methods where other proteins, polypeptides, nucleic acids, antibodies, or binding partners are the targets.

The assay compound can be selected from a library of small molecules, organic compounds which are either synthetic or natural, or mimetic libraries of randomized oligonucleotide-derived or peptide-derived compounds, for example. The compounds of the libraries may contain random chemical modifications, such as acylation, alkylation, esterification, amidation, or other modifications. Ideally, the largest number of separate structural entities will exist in a library that is tested against the agents of the invention for detectable interaction. A variety of other

reagents may be used in the assay, such as buffers, salts, detergents, proteins, protease inhibitors, nuclease inhibitors, antimicrobial agents, or other reagents.

Detecting the interaction between the assay compound and the agent of the invention can be performed via a number of techniques. Fluorescence quenching, specific binding as with avidin-biotin, enzymatic activity, or inhibition of enzymatic activity are examples of the types of techniques used to detect interaction between two molecules. One of skill in the art can devise many specific assays depending on the activity sought. The type of assay used is not crucial to the use of this invention.

10 The screening methods may optionally employ a solid substrate to which one or more assay components are bound. Also, cell-based assays are often used in high throughput screening methods, so that the cell contains or expresses a component of the assay. Numerous permutations are possible.

Each of the activities listed below may be screened for, alone or in combination, in a method to detect an interaction with agents of the invention.

a. Inflammatory or Anti-Inflammatory Activity

Proteins or polypeptides of the present invention may also exhibit inflammatory or anti-inflammatory activity. These activities may relate to a stimulus to cells involved in the inflammatory response, inhibiting or promoting cell-cell interactions (for example, cell adhesion), inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or stimulating or suppressing production of other factors, which more directly inhibit or promote an inflammatory response.

Proteins or polypeptides exhibiting anti-inflammatory activity or antibodies to inflammatory proteins or polypeptides can be used to treat atopic disorders and other inflammatory conditions including: chronic or acute inflammatory conditions, inflammation associated with infection, such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS); ischemia-

reperfusion injury; endotoxin lethality; arthritis; complement-mediated hyperacute rejection; nephritis; cytokine or chemokine-induced lung injury; inflammatory bowel disease; Crohn's disease; or disorders resulting from over-production of cytokines such as TNF or IL-1. Proteins or polypeptides or antibodies of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

b. Cytokine and Cell Growth or Differentiation Activity

A protein or polypeptide of the invention may exhibit cytokine, cell growth promoting or inhibiting, or cell differentiation promoting or inhibiting activity. Many protein factors secreted by immune cells, including cytokines, have exhibited activity in one or more factor dependent cell-based assays. These assays can be used to identify useful activities. The activity of a protein or polypeptide of the invention may be measured by the following methods or others known in the art.

Assays for T-cell or thymocyte proliferation include those described in *Current Protocols in Immunology*, Coligan, *et al.* Eds., Greene Publishing Associates and Wiley-Interscience (1994) (see, especially Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, *Immunologic studies in Humans*); Takai *et al.*, *J. Immunol.* 137:3494-3500 (1986); Bertagnolli *et al.*, *J. Immunol.* 145:1706-1712 (1990); Bertagnolli *et al.*, *Cellular Immunology* 133:327-341 (1991); Bertagnolli, *et al.*, *J. Immunol.* 149:3778-3783 (1992); Bowman *et al.*, *J. Immunol.* 152: 1756-1761 (1994)).

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells, or thymocytes include those described in Polyclonal T cell stimulation, Kruisbeek, and Shevach, In *Current Protocols in Immunology*, Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto (1994); and *Measurement of mouse and human Interferon gamma*, Schreiber, In *Current Protocols in*

Immunology, Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto (1994).

- Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include those described in *Measurement of Human and Murine Interleukin 2 and Interleukin 4*, Bottomly, Davis, and Lipsky In *Current Protocols in Immunology*. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto (1994); deVries *et al.*, *J. Exp. Med.* 173:1205-1211 (1991); Moreau *et al.*, *Nature* 336:690-692 (1988); Greenberger *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 80:2931-2938 (1983); *Measurement of mouse and human interleukin 6*, Nordan, R. In *Current Protocols in Immunology*. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto (1994); Smith *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 83:1857-1861 (1986); *Measurement of human Interleukin 11*, Bennett, *et al.*, In *Current Protocols in Immunology*. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto (1991); *Measurement of mouse and human Interleukin 9*, Ciarletta, *et al.*, In *Current Protocols in Immunology*. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto (1991).

- Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect antigen-presenting cell (APC)-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: *Current Protocols in Immunology*, Coligan, *et al.* eds., Pub. Greene Publishing Associates and Wiley-Interscience (1994)(Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function; Chapter 6, *Cytokines and their cellular receptors*; Chapter 7, *Immunologic studies in Humans*); Weinberger *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 77:6091-6095, 1980; Weinberger *et al.*, *Eur. J. Immun.* 11:405-411, 1981; Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Takai *et al.*, *J. Immunol.* 140:508-512 (1988).

**c. Immunosuppressive, Immune Stimulating, or
Immune Modulating Activity**

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: *Current Protocols in Immunology*, Ed by Coligan, *et al.*, Pub. Greene Publishing Associates and Wiley-Interscience (1994)(Chapter 3, *In Vitro assays for Mouse Lymphocyte Function* 3.1-3.19; Chapter 7, *Immunologic studies in Humans*); Herrmann *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 78:2488-2492 (1981); Herrmann *et al.*, *J. Immunol.* 128:1968-1974 (1982); Handa *et al.*, *J. Immunol.* 135:1564-1572 (1985); Takai *et al.*, *J. Immunol.* 137:3494-3500 (1986); Takai *et al.*, *J. Immunol.* 140:508-512 (1988); Bowman *et al.*, *J. Virology* 61:1992-1998; Bertagnolli *et al.*, *Cellular Immunology* 133:327-341, 1991; Brown *et al.*, *J. Immunol.* 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, *J. Immunol.* 144:3028:3033 (1990); and *Assays for B cell function: In vitro antibody production*, Mond, and Brunswick, In *Current Protocols in Immunology*, Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto (1994).

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: *Current Protocols in Immunology*, Ed by Coligan, *et al.*, Strober, Pub. Greene Publishing Associates and Wiley-Interscience (1994) (Chapter 3, *In Vitro assays for Mouse Lymphocyte Function* 3.1-3.19; Chapter 7, *Immunologic studies in Humans*); Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986; Takai *et al.*, *J. Immunol.* 140:508-512, 1988; Bertagnolli *et al.*, *J. Immunol.* 149:3778-3783 (1992).

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without

limitation, those described in: Guery *et al.*, *J. Immunol.* 134:536-544, 1995; Inaba *et al.*, *Journal of Experimental Medicine* 173:549-559, 1991; Macatonia *et al.*, *J Immunol* 154:5071-5079 (1995); Porgador *et al.*, *Journal of Experimental Medicine* 182:255-260 (1995); Nair *et al.*, *J. Virology* 67:4062-4069 (1993);
5 Huang *et al.*, *Science* 264:961-965 (1994); Macatonia *et al.*, *Journal of Experimental Medicine* 169:1255-1264 (1989); Bhardwaj *et al.*, *Journal of Clinical Investigation* 94:797-807 (1994); and Inaba *et al.*, *Journal of Experimental Medicine* 172:631-640 (1990).

Assays for lymphocyte survival/apoptosis (which will identify, among
10 others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz *et al.*, *Cytometry* 13:795-808 (1992); Gorczyca *et al.*, *Leukemia* 7:659-670 (1993); Gorczyca *et al.*, *Cancer Research* 53:1945-1951 (1993); Itoh *et al.*, *Cell* 66:233-243 (1991); Zacharchuk, *J. Immunol.* 145:4037-4045 (1990);
15 Zamai *et al.*, *Cytometry* 14:891-897 (1993); Gorczyca *et al.*, *International Journal of Oncology* 1:639-648 (1992).

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica *et al.*, *Blood* 84:111-117 (1994); Fine *et al.*, *Cellular Immunology* 155:111-122 (1994); Galy *et al.*, *Blood* 85:2770-2778 (1995); Toki *et al.*, *Proc. Nat. Acad. Sci. (U.S.A.)* 88:7548-7551 (1991).
20

d. Cell Differentiation Activity

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include,
25 without limitation, those described in: Johansson *et al.*, *Cellular Biology* 15:141-151 (1995); Keller *et al.*, *Molecular and Cellular Biology* 13:473-486 (1993); McClanahan *et al.*, 81:2903-2915 (1993).

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: *Methylcellulose colony forming assays*, Freshney, In *Culture of Hematopoietic Cells*, Freshney, *et al.*, eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. (1994); Hirayama *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 89:5907-5911 (1992); Primitive hematopoietic colony forming cells with high proliferate potential, McNiece, and Briddell, In *Culture of Hematopoietic Cells*, Freshney, *et al.*, eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. (1994); Neben *et al.*, *Experimental Hematology* 22:353-359 (1994); Cobblestone area forming cell assay, Ploemacher, In *Culture of Hematopoietic Cells*, Freshney, *et al.*, eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. (1994); and Long term bone marrow cultures in the presence of stromal cells, Spooncer, *et al.*, In *Culture of Hematopoietic Cells*, Freshney, *et al.*, eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. (1994); Long term culture initiating cell assay, Sutherland, In *Culture of Hematopoietic Cells*, R. I. Freshney, *et al.*, eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. (1994).

e. Wound Healing Activity

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); and International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, *Epidermal Wound Healing*, pps. 71-112 (Maibach, and Rovee, eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, *J. Invest. Dermatol* 71:382-84 (1978).

f. Chemotactic Activity

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the controlled orientation or

movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

5 The activity of a protein of the invention may, among other means, be measured by the following methods:

 Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein
10 to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: *Current Protocols in Immunology*, Ed by Coligan *et al.*, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub *et al.*, *J. Clin. Invest.* 95:1370-1376 (1995);
15 Lind *et al.*, *APMIS* 103:140-146 (1995); Muller *et al.*, *Eur. J. Immunol.* 25: 1744-1748; Gruber *et al.*, *J. of Immunol.* 152:5860-5867 (1994); and Johnston *et al.* *J. of Immunol.* 153: 1762-1768 (1994).

g. Receptor/Ligand Interaction Activity

 Proteins of the present invention may also demonstrate activity as
20 receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as
25 selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). These proteins or fragments of the invention, or cells containing them, can be incorporated into an assay to screen for binding to

extracellular matrix proteins, their analogs, or for receptor/ligand interaction to compounds implicated in binding to extracellular matrix proteins. In a particularly preferred embodiment, molecules containing the peptide motif RGD can be used to screen for interaction with the proteins, fragments, or cells of the invention. Various specific assays can be used as a basis for designing the reagents for screening, such as phage attachment assays, panning assays, cell attachment assays, and inhibition of cell attachment/adhesion assays (Pasquelina *et al.*, *J. Cell Biol.* 130:1189-1196 (1995), Koivunen *et al.*, *BioTechnology* 13: 265-270 (1995), Koivunen *et al.*, *Methods Enzym.* 245: 346-369 (1994), and U.S. Patent 5,817,750, all incorporated herein in their entirety). These receptor/ligand interaction assays can also be designed for use with libraries of compounds, such as phage display libraries.

Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. Proteins of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: *Current Protocols in Immunology*, Ed by Coligan, *et al.*, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under Static Conditions 7.28.1-7.28.22); Takai *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 84:6864-6868 (1987); Bierer *et al.*, *J. Exp. Med.* 168:1145-1156 (1988); Rosenstein *et al.*, *J. Exp. Med.* 169:149-160 (1989); Stoltenborg *et al.*, *J. Immunol. Methods* 175:59-68, 1994; Stitt *et al.*, *Cell* 80:661-670 (1995).

2. Methods for Detecting and Manipulating Nucleic Acids

The nucleic acids of the invention can be used directly in numerous methods to identify or detect the presence of specific nucleic acid sequences. As noted above, the nucleic acids of the invention can be used as hybridization probes or PCR probes, or to derive specific hybridization or PCR probes. Agents of the present invention may be labeled with reagents that facilitate detection (e.g., fluorescent labels, Prober *et al.*, *Science* 238: 336-340 (1987), Albarella *et al.*, EP 144914,; chemical labels, Sheldon *et al.*, U.S. Patent 4,582,789, Albarella *et al.*, U.S. Patent 4,563,417; and modified bases, Miyoshi *et al.*, EP 119448)).

Furthermore, the nucleic acids of the invention of variants or fragments thereof can be linked to solid supports. In this way, various microarrays, beads, glass or nylon slides, membranes or other repeatable assay apparatus can be constructed. A non-limiting description of selected methods follows.

a. Microarrays

In one embodiment, the nucleic acids of the invention can be used to monitor expression. A microarray-based method for high-throughput monitoring of gene expression may be utilized to measure activated OA Marker hybridization targets. This 'chip'-based approach involves using microarrays of nucleic acids as specific hybridization targets to quantitatively measure expression of the corresponding genes (Schena *et al.*, *Science* 270:467-470 (1995); Shalon, Ph.D. Thesis, Stanford University (1996). Every nucleotide in a large sequence can be queried at the same time. Hybridization can also be used to efficiently analyze nucleotide sequences.

Several microarray methods have been described. One method compares the sequences to be analyzed by hybridization to a set of oligonucleotides or cDNA molecules representing all possible subsequences (Bains and Smith, *J. Theor. Biol.* 135:303 (1989)). A second method hybridizes the sample to an array of oligonucleotide or cDNA probes. An array consisting of oligonucleotides or

cDNA molecules complementary to subsequences of a target sequence can be used to determine the identity of a target sequence, measure its amount, and detect differences between the target and a reference sequence. Nucleic acid microarrays may also be screened with protein molecules or fragments thereof to determine
5 nucleic acids that specifically bind protein molecules or fragments thereof.

The microarray approach may also be used with polypeptide targets (*see*, U.S. Patent Nos. 5,800,992, 5,445,934; 5,143,854, 5,079,600, 4,923,901). Essentially, polypeptides are synthesized on a substrate (microarray) and these polypeptides can be screened with either protein molecules or fragments thereof
10 nucleic acid molecules in order to screen for either protein molecules or fragments thereof or nucleic acid molecules that specifically bind the target polypeptides (Fodor *et al.*, *Science* 251:767-773 (1991)).

b. Hybridization Assays

Oligonucleotide probes, whose sequences are complementary to that of a
15 portion of the nucleic acids of the invention, such as SEQ NO.:1-82, can be constructed. These probes are then incubated with cell extracts of a patient under conditions sufficient to permit nucleic acid hybridization. The detection of double-stranded probe-mRNA hybrid molecules is indicative of the presence of OA Markers or sequences derived from OA Markers. Thus, such probes may be
20 used to ascertain the level and extent of OA severity or progression or the production of certain proteins. The nucleic acid hybridization may be conducted under quantitative conditions or as a qualitative assay.

c. PCR Assays

A nucleic acid of the invention, such as one of SEQ NO.:1-82 or
25 complements thereof, can be analyzed for use as a PCR probe. A search of databases indicates the presence of regions within that nucleic acid that have high and low regions of identity to other sequences in the database. Ideally, a PCR probe will have high identity with only the sequence from which it is derived. In

that way, only the desired sequence is amplified. Computer generated searches using programs such as MIT Primer3 (Rozen and Skaletsky (1996, 1997, 1998)) or GeneUp (Pesole, *et al.*, *BioTechniques* 25:112-123 (1998)), for example, can be used to identify potential PCR primers.

- 5 The PCR probes or primers can be used in methods such as described in Krzesicki, *et al.*, *Am. J. Respir. Cell Mol. Biol.* 16:693-701 (1997) to identify or detect sequences expressed in osteoarthritis.

d. Ligation and Alternative Amplification Methods and Identification of Polymorphisms

- 10 In one sub-aspect of the invention analysis is conducted to determine the presence and/or identity of polymorphism(s) using one or more of the nucleic acid molecules of the present invention and more specifically one or more of the EST nucleic acid molecule or fragment thereof which are associated with a phenotype, or a predisposition to that phenotype.

- 15 Any of a variety of molecules can be used to identify such polymorphism(s). In one embodiment, one or more of the EST nucleic acid molecules (or a sub-fragment thereof) may be employed as a marker nucleic acid molecule to identify such polymorphism(s). Alternatively, such polymorphisms can be detected through the use of a marker nucleic acid molecule or a marker
20 protein that is genetically linked to (i.e., a polynucleotide that co-segregates with) such polymorphism(s).

- In an alternative embodiment, such polymorphisms can be detected through the use of a marker nucleic acid molecule that is physically linked to such polymorphism(s). For this purpose, marker nucleic acid molecules comprising a
25 nucleotide sequence of a polynucleotide located within 1 mb of the polymorphism(s), and more preferably within 100 kb of the polymorphism(s), and most preferably within 10 kb of the polymorphism(s) can be employed.

The genomes of animals and plants naturally undergo spontaneous mutation in the course of their continuing evolution, creating polymorphisms (Gusella, *Ann. Rev. Biochem.* 55:831-854 (1986)). The variant sequence and the "original" sequence co-exist in the species' population. In some instances, such co-existence is in
5 stable or quasi-stable equilibrium.

A polymorphism is thus said to be "allelic." in that, due to the existence of the polymorphism, some members of a species may have the original sequence (i.e., the original "allele") whereas other members may have the variant sequence (i.e., the variant "allele"). In the simplest case, only one variant sequence may
10 exist, and the polymorphism is thus said to be di-allelic. In other cases, the species' population may contain multiple alleles, and the polymorphism is termed tri-allelic, etc. A single gene may have multiple different unrelated polymorphisms. For example, it may have a di-allelic polymorphism at one site, and a multi-allelic polymorphism at another site.

15 The variation that defines the polymorphism may range from a single nucleotide variation to the insertion or deletion of extended regions within a gene. In some cases, the DNA sequence variations are in regions of the genome that are characterized by short tandem repeats (STRs) that include tandem di- or tri-nucleotide repeated motifs of nucleotides. Polymorphisms characterized by such
20 tandem repeats are referred to as "variable number tandem repeat" ("VNTR") polymorphisms. VNTRs have been used in identity analysis (Weber, U.S. Patent 5,075,217; Armour, *et al.*, *FEBS Lett.* 307:113-115 (1992); Jones, *et al.*, *Eur. J. Haematol.* 39:144-147 (1987); Horn, *et al.*, PCT Patent Application WO91/14003; Jeffreys, European Patent Application 370,719; Jeffreys, U.S. Patent 5,175,082;
25 Jeffreys *et al.*, *Amer. J. Hum. Genet.* 39:11-24 (1986); Jeffreys *et al.*, *Nature* 316:76-79 (1985); Gray *et al.*, *Proc. R. Acad. Soc. Lond.* 243:241-253 (1991); Moore *et al.*, *Genomics* 10:654-660 (1991); Jeffreys *et al.*, *Anim. Genet.* 18:1-15

(1987); Hillel *et al.*, *Anim. Genet.* 20:145-155 (1989); Hillel *et al.*, *Genet.* 124:783-789 (1990)).

The detection of polymorphic sites in a sample of DNA may be facilitated through the use of nucleic acid amplification methods. Such methods specifically
5 increase the concentration of polynucleotides that span the polymorphic site, or include that site and sequences located either distal or proximal to it. Such amplified molecules can be readily detected by gel electrophoresis or other means.

The most preferred method of achieving such amplification employs the polymerase chain reaction ("PCR")(See above). In lieu of PCR, alternative
10 amplification methods, such as the "Ligase Chain Reaction" ("LCR") may be used (Barany, *Proc. Natl. Acad. Sci. (U.S.A.)* 88:189-193 (1991). LCR uses two pairs of oligonucleotide probes to exponentially amplify a specific target. The sequences of each pair of oligonucleotides is selected to permit the pair to hybridize to abutting sequences of the same strand of the target. Such
15 hybridization forms a substrate for a template-dependent ligase. As with PCR, the resulting products thus serve as a template in subsequent cycles and an exponential amplification of the desired sequence is obtained.

LCR can be performed with oligonucleotides having the proximal and distal sequences of the same strand of a polymorphic site. In one embodiment,
20 either oligonucleotide will be designed to include the actual polymorphic site of the polymorphism. In such an embodiment, the reaction conditions are selected such that the oligonucleotides can be ligated together only if the target molecule either contains or lacks the specific nucleotide that is complementary to the polymorphic site present on the oligonucleotide. Alternatively, the
25 oligonucleotides may be selected such that they do not include the polymorphic site (see, Segev, PCT Application WO 90/01069).

The "Oligonucleotide Ligation Assay" ("OLA") may, alternatively be employed (Landegren, *et al.*, *Science* 241:1077-1080 (1988)). The OLA protocol

uses two oligonucleotides which are designed to be capable of hybridizing to abutting sequences of a single strand of a target. OLA, like LCR, is particularly suited for the detection of point mutations. Unlike LCR, however, OLA results in "linear" rather than exponential amplification of the target sequence.

5 Nickerson, *et al.* have described a nucleic acid detection assay that combines attributes of PCR and OLA (Nickerson, *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 87:8923-8927 (1990). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA. In addition to requiring multiple, and separate, processing steps, one problem
10 associated with such combinations is that they inherit all of the problems associated with PCR and OLA.

Schemes based on ligation of two (or more) oligonucleotides in the presence of nucleic acid having the sequence of the resulting "di-oligonucleotide," thereby amplifying the di-oligonucleotide, are also known (Wu, *et al.*, *Genomics*
15 4:560 (1989)), and may be readily adapted to the purposes of the present invention.

Other known nucleic acid amplification procedures, such as allele-specific oligomers, branched DNA technology, transcription-based amplification systems, or isothermal amplification methods may also be used to amplify and analyze such
20 polymorphisms (Malek, *et al.*, U.S. Patent 5,130,238; Davey, *et al.*, European Patent Application 329,822; Schuster *et al.*, U.S. Patent 5,169,766; Miller, *et al.*, PCT appln. WO 89/06700; Kwoh, *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 86:1173 (1989); Gingeras, *et al.*, PCT application WO 88/10315; Walker, *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.)* 89:392-396 (1992)). Any of the foregoing nucleic acid
25 amplification methods could be used.

The identification of a polymorphism in a gene, fragment or cellular sequence derived from the nucleic acids of the invention can be determined in a variety of ways. By correlating the presence or absence of atopic disease, for

example, in an individual with the presence or absence of a polymorphism, it is possible to diagnose the predisposition of a patient to osteoarthritis-related disorders. If a polymorphism creates or destroys a restriction endonuclease cleavage site, or if it results in the loss or insertion of DNA (e.g., a VNTR polymorphism), it will alter the size or profile of the DNA fragments that are generated by digestion with that restriction endonuclease. As such, individuals that possess a variant sequence can be distinguished from those having the original sequence by restriction fragment analysis. Polymorphisms that can be identified in this manner are termed "restriction fragment length polymorphisms" ("RFLPs"). RFLPs have been widely used in human and animal genetic analyses (Glassberg, UK Patent Application 2135774; Skolnick, *et al.*, *Cytogen. Cell Genet.* 32:58-67 (1982); Botstein, *et al.*, *Ann. J. Hum. Genet.* 32:314-331 (1980); Fischer, *et al.* (PCT Application WO90/13668); Uhlen, PCT Application WO90/11369).

Other types of polymorphisms include single nucleotide polymorphisms (SNPs) that are single base changes in genomic DNA sequence. They generally occur at greater frequency than other markers and are spaced with a greater uniformity throughout a genome than other reported forms of polymorphism. The greater frequency and uniformity of SNPs means that there is greater probability that such a polymorphism will be found near or in a genetic locus of interest than would be the case for other polymorphisms. SNPs are located in protein-coding regions and noncoding regions of a genome. Some of these SNPs may result in defective or variant protein expression (e.g., as a result of mutations or defective splicing). Analysis (genotyping) of characterized SNPs can require only a plus/minus assay rather than a lengthy measurement, permitting easier automation.

Modifications can be naturally provided or deliberately engineered into the nucleic acids, proteins, and polypeptides of the invention to generate variants. For example, modifications in the peptide or DNA sequences can be made by those

skilled in the art using known techniques, such as site-directed mutagenesis. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of one or more selected amino acid residues. For example, one or more cysteine residues may be deleted or replaced
5 with another amino acid to alter the conformation of the molecule. Additional cysteine residues can also be added as a substitute at sites to promote disulfide bonding and increase stability. Techniques for identifying the sites for alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art. Techniques for making alterations, substitutions, replacements, insertions
10 or deletions (see, e.g., U.S. Pat. No. 4,518,584) are also well known in the art. Preferably, any modification of a protein, polypeptide, or nucleic acid of the invention will retain at least one of the structural or functional attributes of the molecule.

SNPs can be characterized using any of a variety of methods. Such
15 methods include the direct or indirect sequencing of the site, the use of restriction enzymes (Botstein *et al.*, *Am. J. Hum. Genet.* 32:314-331 (1980)); Konieczny and Ausubel, *Plant J.* 4:403-410 (1993), enzymatic and chemical mismatch assays (Myers *et al.*, *Nature* 313:495-498 (1985), allele-specific PCR (Newton *et al.*, *Nucl. Acids Res.* 17:2503-2516 (1989)); Wu *et al.*, *Proc. Natl. Acad. Sci. USA*
20 86:2757-2760 (1989)), ligase chain reaction (Barany, *Proc. Natl. Acad. Sci. USA* 88:189-193 (1991)), single-strand conformation polymorphism analysis (Labruno *et al.*, *Am. J. Hum. Genet.* 48: 1115-1120 (1991)), primer-directed nucleotide incorporation assays (Kuppuswami *et al.*, *Proc. Natl. Acad. Sci. USA* 88:1143-1147 (1991)), dideoxy fingerprinting (Sarkar *et al.*, *Genomics* 13:441-443 (1992)),
25 solid-phase ELISA-based oligonucleotide ligation assays (Nikiforov *et al.*, *Nucl. Acids Res.* 22:4167-4175 (1994)), oligonucleotide fluorescence-quenching assays (Livak *et al.*, *PCR Methods Appl.* 4:357-362 (1995a)), 5'-nuclease allele-specific hybridization TaqMan assay (Livak *et al.*, *Nature Genet.* 9:341-342 (1995)),

template-directed dye-terminator incorporation (TDI) assay (Chen and Kwok, *Nucl. Acids Res.* 25:347-353 (1997)), allele-specific molecular beacon assay (Tyagi *et al.*, *Nature Biotech.* 16: 49-53 (1998)), PinPoint assay (Haff and Smirnov, *Genome Res.* 7: 378-388 (1997)), and dCAPS analysis (Neff *et al.*, *Plant J.* 14:387-392 (1998)).

SNPs can be observed by examining sequences of overlapping clones in the BAC library according to the method described by Taillon-Miller *et al.* *Genome Res.* 8:748-754 (1998)). SNPs can also be observed by screening the BAC library of the present invention by colony or plaque hybridization with a labeled probe containing SNP markers; isolating positive clones and sequencing the inserts of the positive clones; suitable primers flanking the SNP markers.

Polymorphisms can also be identified by Single Strand Conformation Polymorphism (SSCP) analysis. SSCP is a method capable of identifying most sequence variations in a single strand of DNA, typically between 150 and 250 nucleotides in length (Elles, *Methods in Molecular Medicine: Molecular Diagnosis of Genetic Diseases*, Humana Press (1996); Orita *et al.*, *Genomics* 5:874-879 (1989)). Under denaturing conditions a single strand of DNA will adopt a conformation that is uniquely dependent on its sequence conformation. This conformation usually will be different, even if only a single base is changed. Most conformations have been reported to alter the physical configuration or size sufficiently to be detectable by electrophoresis. A number of protocols have been described for SSCP including, but not limited to, Lee *et al.*, *Anal. Biochem.* 205:289-293 (1992); Suzuki *et al.*, *Anal. Biochem.* 192:82-84 (1991); Lo *et al.*, *Nucleic Acids Research* 20: 1005-1009 (1992); and Sarkar *et al.*, *Genomics* 13:441-443 (1992). It is understood that one or more of the nucleic acids of the present invention, may be utilized as markers or probes to detect polymorphisms by SSCP analysis.

Polymorphisms may also be found using a DNA fingerprinting technique called amplified fragment length polymorphism (AFLP), which is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA to profile that DNA (Vos *et al.*, *Nucleic Acids Res.* 23:4407-4414
5 (1995)). This method allows for the specific co-amplification of high numbers of restriction fragments, which can be visualized by PCR without knowledge of the nucleic acid sequence.

AFLP employs basically three steps. Initially, a sample of genomic DNA is cut with restriction enzymes and oligonucleotide adapters are ligated to the
10 restriction fragments of the DNA. The restriction fragments are then amplified using PCR by using the adapter and restriction sequence as target sites for primer annealing. The selective amplification is achieved by the use of primers that extend into the restriction fragments, amplifying only those fragments in which the primer extensions match the nucleotide flanking the restriction sites. These
15 amplified fragments are then visualized on a denaturing polyacrylamide gel.

AFLP analysis has been performed on *Salix* (Beismann *et al.*, *Mol. Ecol.* 6:989-993 (1997)), *Acinetobacter* (Janssen *et al.*, *Int. J. Syst. Bacteriol.* 47:1179-1187 (1997)), *Aeromonas popoffi* (Huys *et al.*, *Int. J. Syst. Bacteriol.* 47:1165-1171 (1997)), *Phytophthora infestans* (Van der Lee *et al.*, *Fungal Genet. Biol.* 20 21:278-291 (1997)), *Bacillus anthracis* (Keim *et al.*, *J. Bacteriol.* 179:818-824 (1997)), *Astragalus cremonophylax* (Travis *et al.*, *Mol. Ecol.* 5:735-745 (1996)), *Arabidopsis* (Cnops *et al.*, *Mol. Gen. Genet.* 253:32-41 (1996)), *Escherichia coli* (Lin *et al.*, *Nucleic Acids Res.* 24:3649-3650 (1996)), *Aeromonas* (Huys *et al.*, *Int. J. Syst. Bacteriol.* 46:572-580 (1996)), nematode (Folkertsma *et al.*, *Mol. Plant*
25 *Microbe Interact.* 9:47-54 (1996)), and human (Latorra *et al.*, *PCR Methods Appl.* 3:351-358 (1994)). AFLP analysis has also been used for fingerprinting mRNA (Money *et al.*, *Nucleic Acids Res.* 24:2616-2617 (1996), the entirety of which is herein incorporated by reference; Bachem *et al.*, *Plant J.* 9:745-753 (1996)). It is

understood that one or more of the nucleic acids of the present invention, may be utilized as markers or probes to detect polymorphisms by AFLP analysis or for fingerprinting RNA.

The polymorphism obtained by these approaches can then be cloned to
5 identify the mutation at the coding region, which alters the protein's structure or the regulatory region of the gene that affects its expression level. Changes involving promoter interactions with other regulatory proteins can be identified by, for example, gel shift.

In accordance with an embodiment of the invention, a sample DNA is
10 obtained from a patient's cells. In a preferred embodiment, the DNA sample is obtained from the patient's blood. However, any source of DNA may be used. The DNA may be subjected to interrogation to determine the presence or absence of a polymorphism.

15 3. Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., antisense, polypeptides, peptidomimetics, small molecules or other drugs) which bind to nucleic acid molecules, polypeptides or proteins described herein or
20 have a stimulatory or inhibitory effect on, for example, expression or activity of the nucleic acid molecules, polypeptides or proteins of the invention.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of protein or polypeptide described herein or biologically active portion thereof. The test compounds of the
25 present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library

method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to polypeptide libraries, while the other four approaches are applicable to polypeptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K. S. (1997) *Anticancer Drug Des.*

5 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.*, 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. U.S.A.*, 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.*, 37:2678; Cho *et al.* (1993) *Science*, 261:1303; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.*, 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.*, 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.*, 37:1233.

Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques*, 13:412-421), or on beads (Lam(1991) *Nature*, 354:82-84), chips (Fodor (1993) *Nature*, 364:555-556), bacteria(Ladner U.S. Pat. No. 5,223,409), spores (Ladner U.S. Pat. No. '409), plasmids (Cull *et al.*(1992) *Proc. Natl. Acad. Sci. U.S.A.*, 89:1865-1869) or on phage (Scott and Smith (1990) *Science*, 249:386-390); (Devlin (1990) *Science*, 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.*, 97:6378-6382); (Felici (1991) *J. Mol. Biol.*, 222:301-310); (Ladner *supra*).

In one embodiment, an assay is a cell-based assay in which a cell which expresses an encoded protein which is a cell surface receptor is contacted with a test compound and the ability of the test compound to bind to the receptor is determined. The cell, for example, can be of mammalian origin, such as from cartilage tissue or bone. Determining the ability of the test compound to bind to the receptor can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the receptor can be determined by detecting the labeled with ¹²⁵I, ³⁵S,

.sup.14 C, or .sup.3 H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected
5 by determination of conversion of an appropriate substrate to product.

It is also within the scope of this invention to determine the ability of a test compound to interact with the receptor without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a test compound with the receptor without the labeling of either the
10 test compound or the receptor. McConnell, H. M. *et al.* (1992) Science, 257:1906-1912. A "microphysiometer" (e.g., Cytosensor.TM.) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between ligand and receptor.

15 In one embodiment, the assay comprises contacting a cell which expresses an encoded protein described herein on the cell surface (e.g., a receptor) with a receptor ligand or biologically-active portion thereof, to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the receptor, wherein determining the ability of
20 the test compound to interact with the receptor comprises determining the ability of the test compound to preferentially bind to the receptor as compared to the ability of the ligand, or a biologically active portion thereof, to bind to the receptor.

In another embodiment, an assay is a cell-based assay comprising
25 contacting a cell expressing a particular target molecule described herein with a test compound and determining the ability of the test compound to modulate or alter (e.g. stimulate or inhibit) the activity of the target molecule. Determining the ability of the test compound to modulate the activity of the target molecule can be

accomplished, for example, by determining the ability of a known ligand to bind to or interact with the target molecule.

Determining the ability of the known ligand to bind to or interact with the target molecule can be accomplished by one of the methods described above for
5 determining direct binding. In a preferred embodiment, determining the ability of the known ligand to bind to or interact with the target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (e.g., intracellular Ca^{2+} ,
10 diacylglycerol, IP₃, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a target-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, e.g., luciferase), or detecting a cellular response, for example, development, differentiation or rate of proliferation.

15 In yet another embodiment, an assay of the present invention is a cell-free assay in which protein of the invention or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the protein or biologically active portion thereof is determined. Binding of the test compound to the protein can be determined either directly or indirectly as
20 described above. In one embodiment, the assay includes contacting the protein or biologically active portion thereof with a known compound which binds the protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the protein. Determining the ability of the test compound to interact with the protein
25 comprises determining the ability of the test compound to preferentially bind to the protein or biologically active portion thereof as compared to the known compound.

In another embodiment, the assay is a cell-free assay in which a protein of the invention or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate or alter (e.g., stimulate or inhibit) the activity of the protein or biologically active portion thereof is determined. Determining the ability of the test compound to modulate the activity of the protein can be accomplished, for example, by determining the ability of the protein to bind to a known target molecule by one of the methods described above for determining direct binding. Determining the ability of the protein to bind to a target molecule can also be accomplished using a technology such as real-time Bimolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.*, 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.*, 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore.TM.). Changes in the optical phenomenon surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

In an alternative embodiment, determining the ability of the test compound to modulate the activity of a protein of the invention can be accomplished by determining the ability of the protein to further modulate the activity of a target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

In yet another embodiment, the cell-free assay involves contacting a protein of the invention or biologically active portion thereof with a known compound which binds the protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the protein, wherein determining the ability of the test compound to interact with the protein comprises determining the ability of the protein to preferentially bind to or modulate the activity of a target molecule.

The cell-free assays of the present invention are amenable to use of both soluble and/or membrane-bound forms of isolated proteins. In the case of cell-free assays in which a membrane-bound form of an isolated protein is used it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the isolated protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton.RTM.X-100, Triton.RTM.X-114, Thesit.RTM., Isotridecypoly(ethylene glycol ether).sub.n,3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either the protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to the protein, or interaction of the protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, Mo.) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or protein of the invention, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate

wells are washed to remove any unbound components. the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of binding or activity determined using standard techniques.

5 Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a protein of the invention or a target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated protein of the invention or target molecules can be prepared from biotin-NHS(N-hydroxysuccinimide) using techniques well known
10 in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with a protein of the invention or target molecules, but which do not interfere with binding of the protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or
15 protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the protein or target molecule.

20 In another embodiment, modulators of expression of nucleic acid molecules of the invention are identified in a method wherein a cell is contacted with a candidate compound and the expression of appropriate mRNA or protein in the cell is determined. The level of expression of appropriate mRNA or protein in the presence of the candidate compound is compared to the level of expression of
25 mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of expression based on this comparison. For example, when expression of mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than

in its absence, the candidate compound is identified as a stimulator or enhancer of the mRNA or protein expression. Alternatively, when expression of the mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of the mRNA or protein expression. The level of mRNA or protein expression in the cells can be determined by methods described herein for detecting mRNA or protein.

In yet another aspect of the invention, the proteins of the invention can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Pat. No. 5,283,317; Zervos *et al.* (1993) *Cell*, 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.*, 268:12046-12054; Bartel *et al.* (1993) *Biotechniques*, 14:920-924; Iwabuchi *et al.* (1993) *Oncogene*, 8:1693-1696; and Brent W094/10300), to identify other proteins (captured proteins) which bind to or interact with the proteins of the invention and modulate their activity. Such captured proteins are also likely to be involved in the propagation of signals by the proteins of the invention as, for example, downstream elements of a protein-mediated signaling pathway. Alternatively, such captured proteins are likely to be cell-surface molecules associated with non-protein-expressing cells, wherein such captured proteins are involved in signal transduction.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a protein of the invention is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, in vivo, forming an protein-dependent

complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be
5 detected, and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the protein of the invention.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to
10 further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a modulating agent, an antisense nucleic acid molecule, a specific antibody, or a protein-binding partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described
15 herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

Animal models appropriate for the study of the progression of OA include
20 the anterior cruciate ligament dog model (Rubin *et al.*, *Clin. Orthop.* 113:212 (1975); Johnson, *Exp. Pathol.* 30:209 (1986); Pelletier *et al.*, *Osteoarthritis Cartilage* 7:416 (1999)), the rabbit partial meniscectomy model (Colombo *et al.*, *Arthritis Rheum.* 26:875 (1983); Butler *et al.*, *Arthritis Rheum.* 26:1380 (1983)), and the guinea pig spontaneous model (Watson *et al.*, *Osteoarthritis Cartilage*
25 4:197 (1996); Jiminez *et al.*, *Lab Anim. Sci.* 47:598 (1997)).

4. Detection Assays

Portions or fragments of the nucleotide sequences identified herein (and the corresponding complete gene sequences) can be used as polynucleotide

reagents. For example, the sequences can be used to map their respective genes on a chromosome and thus locate gene regions associated with genetic disease.

a. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been
5 isolated, this sequence can be used to map the location of the gene on a
chromosome. This process is called chromosome mapping. Accordingly, portions
or fragments of the nucleic acid molecules, described herein, can be used to map
the location of the corresponding genes on a chromosome. The mapping of the
sequences to chromosomes is an important first step in correlating these sequences
10 with genes associated with disease.

Briefly, genes can be mapped to chromosomes by preparing PCR primers
(preferably 15-30 bp in length) from the nucleic acid molecules described herein.
Computer analysis of the sequences can be used to predict primers that do not
span more than one exon in the genomic DNA, thus complicating the
15 amplification process. These primers can then be used for PCR screening of
somatic cell hybrids containing individual human chromosomes. Only those
hybrids containing the human gene corresponding to the appropriate nucleotide
sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different
20 mammals (e.g., human and mouse cells). As hybrids of human and mouse cells
grow and divide, they gradually lose human chromosomes in random order, but
retain the mouse chromosomes. By using media in which mouse cells cannot
grow, because they lack a particular enzyme, but human cells can, the one human
chromosome that contains the gene encoding the needed enzyme, will be retained.
25 By using various media, panels of hybrid cell lines can be established. Each cell
line in a panel contains either a single human chromosome or a small number of
human chromosomes, and a full set of mouse chromosomes, allowing easy
mapping of individual genes to specific human chromosomes. (D'Eustachio P. *et*

al. (1983) *Science*, 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the nucleic acid molecules of the invention to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a specified sequence to its chromosome include in situ hybridization (described in Fan, Y *et al.* (1990) *PNAS*, 97:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence in situ hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical such as colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. for a review of this technique, see Verma *et al.*, *Human Chromosomes: A Manual of Basic Techniques* (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can

be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, *Medelian Inheritance in Man*, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, *J et al.* (1987) *Nature*, 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with a specified gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

25 5. Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically.

Accordingly, one aspect of the present invention relates to diagnostic assays for determining protein and/or nucleic acid expression as well as activity of proteins of the invention, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with activity or expression of proteins or nucleic acids of the invention.

For example, mutations in a specified gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with expression or activity of nucleic acid molecules or proteins of the invention.

Another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of proteins of the invention in clinical trials.

These and other agents are described in further detail in the following sections.

a. Diagnostic Assays

An exemplary method for detecting the presence or absence of proteins or nucleic acids of the invention in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting the protein, or nucleic acid (e.g., mRNA, genomic DNA) that encodes the protein, such that the presence of the protein or nucleic acid is detected in the biological sample. A preferred agent for detecting mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic DNA sequences described herein. The nucleic

acid probe can be, for example, a full-length nucleic acid, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to appropriate mRNA or genomic DNA. Other suitable probes for use in the

5 diagnostic assays of the invention are described herein.

A preferred agent for detecting proteins of the invention is an antibody capable of binding to the protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody or a fragment thereof (e.g., Fab or F(ab₂)₁) can be used. The term

10 "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently

15 labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect mRNA, protein, or

20 genomic DNA of the invention in a biological sample in vitro as well as in vivo. For example, in vitro techniques for detection of mRNA include Northern hybridizations and in situ hybridizations. In vitro techniques for detection of protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. In vitro techniques for detection

25 of genomic DNA include Southern hybridizations. Furthermore, in vivo techniques for detection of protein include introducing into a subject a labeled anti-protein antibody. For example, the antibody can be labeled with a radioactive

marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample or biopsy isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting protein, mRNA, or genomic DNA of the invention, such that the presence of protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of protein, mRNA or genomic DNA in the control sample with the presence of protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of proteins or nucleic acid molecules of the invention in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting protein or mRNA in a biological sample; means for determining the amount of in the sample; and means for comparing the amount of in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect protein or nucleic acid.

b. Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant expression or activity of proteins and nucleic acid molecules of the invention. For example, the assays described herein, such as the preceding diagnostic assays or the following assays can be utilized to identify a subject

having or at risk of developing a disorder associated with protein or nucleic acid expression or activity such as OA. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing OA. Thus, the present invention provides a method for identifying a disease or disorder associated with

5 aberrant expression or activity of proteins or nucleic acid molecules of the invention, in which a test sample is obtained from a subject and protein or nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant expression or activity of the protein or nucleic

10 acid sequence of the invention. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., synovial fluid), cell sample, or tissue (e.g., cartilage tissue).

Furthermore, the prognostic assays described herein can be used to

15 determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, polypeptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant expression or activity of a protein or nucleic acid molecule of the invention. For example, such methods can be used to determine whether a subject can be

20 effectively treated with an agent for a disorder, such as OA. Alternatively, such methods can be used to determine whether a subject can be effectively treated with an agent for OA. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant expression or activity of a protein or nucleic acid of the present

25 invention, in which a test sample is obtained and protein or nucleic acid expression or activity is detected (e.g., wherein the abundance of particular protein or nucleic acid expression or activity is diagnostic for a subject that can be

administered the agent to treat a disorder associated with aberrant expression or activity.)

The methods of the invention can also be used to detect genetic alterations in genes or nucleic acid molecules of the present invention, thereby determining if a subject with the altered gene is at risk for OA. In one embodiment, the method includes detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a particular protein, or the mis-expression of the gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of (1) a deletion of one or more nucleotides; (2) an addition of one or more nucleotides; (3) a substitution of one or more nucleotides, (4) a chromosomal rearrangement; (5) an alteration in the level of a messenger RNA transcript; (6) aberrant modification, such as of the methylation pattern of the genomic DNA; (7) the presence of a non-wild type splicing pattern of a messenger RNA transcript; (8) a non-wild type level; (9) allelic loss; and (10) inappropriate post-translational modification. As described herein, there are a large number of assay techniques known in the art which can be used for detecting alterations in a particular gene. A preferred biological sample is a tissue or serum sample isolated by conventional means from a subject.

In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202), such as an anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.* (1988) *Science*, 241:1077-1080; and Nakazawa *et al.* (1994) *PNAS*, 91:360-364), the latter of which can be particularly useful for detecting point mutations (see Abravaya *et al.* (1995) *Nucleic Acids Res.*, 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample

with one or more primers which specifically hybridize to the gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a
5 control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli, J. C. *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, 87:1874-
10 1878), transcriptional amplification system (Kwoh, D. Y. *et al.*, (1989) *Proc. Natl. Acad. Sci. USA*, 86:1173-1177), Q-Beta Replicase (Lizardi, P. M. *et al.*, (1988) *Bio/Technology*, 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the
15 detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a given gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested
20 with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for sample, U.S. Pat. No. 5,498,531) can be used to score for the presence of specific mutations by
25 development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotide probes (Cronin, M. T. *et al.*

(1996) *Human Mutation*, 7:244-255; Kozal, M. J. *et al.* (1996) *Nature Medicine*, 2:753-759). For example, genetic mutations can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M. T. *et al.* supra. Briefly, a first hybridization array of probes can be used to scan through
5 long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations
10 detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the gene and detect mutations by
15 comparing the sequence of the gene from the sample with the corresponding wild-type (control) gene sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert ((1997) *PNAS*, 74:560) or Sanger ((1977) *PNAS*, 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic
20 assays ((1995) *Biotechniques*, 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.* (1996) *Adv. Chromatogr.*, 36:127-162; and Griffin *et al.* (1993) *Appl. Biochem. Biotechnol.*, 38:147-159).

Other methods for detecting mutations include methods in which
25 protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science*, 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes formed by hybridizing (labeled) RNA or DNA containing the wild-type sequence

with potentially mutant RNA or DNA obtained from a tissue sample. The double-standard duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to base pair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with
5 Rnase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digest the mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example Cotton *et al.* (1988) *Proc. Natl. Acad. Sci. USA*, 85:4397; Saleeba *et al.* (1992) *Methods Enzymol.*, 217:286-295. In a
10 preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair", enzymes) in defined systems for detecting and mapping point mutations in cDNAs obtained from samples of cells. For example,
15 the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis*, 15:1657-1662). According to an exemplary embodiment, a probe based on an nucleotide sequence of the invention is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch
20 repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Pat. No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic
25 mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc. Natl. Acad. Sci. USA*, 86:2766, see also Cotton (1993) *Mutat Res*, 285:125-144; and Hayashi (1992) *Genet Anal. Tech. Appl.*, 9:73-79). Single-stranded DNA fragments of sample and control nucleic acids will be denatured and allowed to

renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by
5 using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet.*, 7:5).

10 In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature*, 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a
15 GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys. Chem.*, 265:12753).

Examples of other techniques for detecting point mutations include, but are
20 not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.* (1986) *Nature*, 324:163); Saiki *et al.* (1989) *Proc. Natl. Acad. Sci.*
25 *USA*, 86:6320). Such allele-specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech*, 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes*, 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci. USA*, 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a gene of the present invention. Any cell type or tissue in which the gene is expressed may be utilized in the prognostic assays described herein.

c. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of nucleic acid molecules or proteins of the present invention (e.g., modulation of cellular signal transduction, regulation of gene transcription in a cell involved in development or differentiation, regulation of

cellular proliferation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase gene expression, protein levels, or upregulate protein activity, can be monitored in clinical trials of subjects exhibiting decreased gene expression, protein levels, or downregulated protein activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease gene expression, protein levels, or downregulate protein activity, can be monitored in clinical trials of subjects exhibiting increased gene expression, protein levels, or upregulated protein activity. In such clinical trials, the expression or activity of the specified gene and, preferably, other genes that have been implicated in, for example, a proliferative disorder can be used as a "read out" or markers of the phenotype of a particular cell.

For example, and not by way of limitation, genes that are modulated in cells by treatment with an agent (e.g., compound, drug or small molecule) which modulates protein activity (e.g., identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on OA, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of the specified gene and other genes implicated in OA. The levels of gene expression (i.e., a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of the specified gene or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (e.g., an

agonist, antagonist, peptidomimetic, protein, polypeptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a specified protein, mRNA, or genomic DNA of the invention in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the protein, mRNA, or genomic DNA in the pre-administration sample with the protein, mRNA, or genomic DNA in the post-administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of the protein or nucleic acid molecule to higher levels than detected, i.e., to increase effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease effectiveness of the agent. According to such an embodiment, protein or nucleic acid expression or activity may be used as an indicator of the effectiveness of an agent, even in the absence of an observable phenotypic response.

20

6. Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant expression or activity of proteins or nucleic acids of the invention. With regard to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical

development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (e.g., a patient's "drug response phenotype", or "drug response genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with the molecules of the present invention or modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug related side effects.

10

a. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with aberrant expression or activity of genes or proteins of the present invention, by administering to the subject an agent which modulates expression or at least one activity of a gene or protein of the invention. Subjects at risk for a disease which is caused or contributed to by aberrant gene expression or protein activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of aberrancy, for example, an agonist or antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the present invention are further discussed in the following subsections.

b. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating expression or activity of genes or proteins of the invention for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of the specified protein associated with the cell. An agent that modulates protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a protein described herein, a polypeptide, a peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more protein activities. Examples of such stimulatory agents include active protein as well as a nucleic acid molecule encoding the protein that has been introduced into the cell. In another embodiment, the agent inhibits one or more protein activities. Examples of such inhibitory agents include antisense nucleic acid molecules and anti-protein antibodies. These modulatory methods can be performed in vitro (e.g., by culturing the cell with the agent) or, alternatively, in vivo (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a protein or nucleic acid molecule of the invention. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) expression or activity of a gene or protein of the invention. In another embodiment, the method involves administering a protein or nucleic acid molecule of the invention as therapy to compensate for reduced or aberrant expression or activity of the protein or nucleic acid molecule.

Stimulation of protein activity is desirable in situations in which the protein is abnormally downregulated and/or in which increased protein activity is likely to have a beneficial effect. Likewise, inhibition of protein activity is

desirable in situations in which the protein is abnormally upregulated and/or in which decreased protein activity is likely to have a beneficial effect.

c. Pharmacogenomics

5 The molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on the protein activity (e.g., gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (e.g., proliferative or developmental disorders) associated with aberrant protein
10 activity. In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active
15 drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a molecule of the invention or modulator thereof, as well as tailoring the dosage and/or therapeutic regimen of treatment with such a molecule or modulator.

Pharmacogenomics deals with clinically significant hereditary variations in
20 the response to drugs due to altered drug disposition and abnormal action in affected persons. See e.g., Eichelbaum, M., *Clin Exp Pharmacol. Physiol.*, (1996) 23(10-11):983-985 and Linder, M. W., *Clin. Chem.* (1997) 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body
25 (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency

(G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (e.g., a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants). Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1,000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug's target is known (e.g., a protein or a receptor of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2(NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (e.g., a molecule or modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment of an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions

or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a molecule or modulator of the invention, such as a modulator identified by one of the exemplary screening assays described herein.

- 5 The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference.

Examples

The following examples will illustrate the invention in greater detail, although it will be understood that the invention is not limited to these specific
10 examples. Various other examples will be apparent to the person skilled in the art after reading the present disclosure without departing from the spirit and scope of the invention. It is intended that all such other examples be included within the scope of the appended claims.

Example 1

15 Identification of Initial ESTs

Highly selective targets for osteoarthritic tissue were identified from
osteoarthritic tissue to a global database of ESTs selected from libraries made
from normal human tissue. Three EST libraries of OA cartilage and synovium
from 6 different donors (five cartilage donors and one synovium donor) were
20 compared to two EST libraries representing cartilage and synovium from 5 donors
(four cartilage donors and one synovium donor) without OA to identify those
ESTs which are representative of genes upregulated as a result of OA. The data
mining effort used sequence comparison techniques based on BLAST comparison
of individual ESTs to evaluate which ESTs were preferentially observed in the
25 target libraries versus control and/or normal libraries. The resulting EST
sequences represent osteoarthritis tissue-derived nucleic acids of the invention and
can be used to create additional osteoarthritis tissue-derived nucleic acids, proteins,

and polypeptides of the invention. The initial ESTs identified from osteoarthritis tissue libraries (Table 1) were run through bioinformatic computer analysis for the presence of specific structural features, and were given the best possible annotation. Table 1 correlates that information. The sequences which were found
5 to be secreted proteins seen at a frequency which statistically correlates with OA disease presence appear in Table 2.

Table 1. Nucleic acid molecules expressed in osteoarthritis tissue with potential uses for the development of human therapeutics and diagnostics.

10 **Table 2:** Preferred candidate markers of OA Example 2

Example 2

Strategy for obtaining FL clones

The sequences, or fragments of them, disclosed can be used to directly obtain full length clones from cDNA libraries and genomic clones from genomic
15 libraries. A number of methods have been described to obtain cDNA and genomic clones, including 5' RACE (for example, using the Marathon cDNA Amplification Kit from Clontech, Inc.), Genetrapp (Life Technologies, Inc.), colony hybridization (Sambrook *et al.*, 1989, *Molecular Cloning, A Laboratory Manual* 2nd Ed., Cold Spring Harbor Press, Cold Spring Harbor, New York) and
20 array hybridization. One skilled in the art can refer to general reference texts for detailed descriptions of known techniques or equivalent techniques. These texts include *Current Protocols in Molecular Biology* (Ausubel, *et al.*, eds., John Wiley & Sons, N.Y. (1989), and supplements through September 1998), *Molecular Cloning, A Laboratory Manual* (Sambrook *et al.*, 2nd Ed., Cold Spring Harbor
25 Press, Cold Spring Harbor, New York (1989)), and *Current Protocols in Immunology* (Coligan, ed., John Wiley and Sons, Toronto (1994)), for example, each of which are specifically incorporated by reference in their entirety.

A preferred method is array hybridization. A first step in array hybridization is to obtain a high quality library with high complexity and a high proportion of full length or high molecular weight clones. A cDNA library can be purchased from a number of commercial sources or a new library prepared from mRNA derived from tissue known or suspected to express the gene. Details on library construction can be found in Sambrook *et al.* (*Molecular Cloning, A Laboratory Manual* 2nd Ed., Cold Spring Harbor Press, Cold Spring Harbor, New York). Similarly, genomic libraries may be purchased from commercial sources.

A plasmid cDNA library constructed from human OA tissue -derived mRNA is plated on agar and a picking robot (for example, a 'Q' BOT from Genetix, Inc.) is used to pick individual colonies into 96 well plates containing LB medium. The isolated *E. coli* cells are grown overnight at 30°C and placed at 4°C until ready for spotting. The *E. coli* are then robotically spotted in high density grids on nitrocellulose membranes overlaying solid agar medium. Preferably, each colony is double spotted at two different locations. The colonies are allowed to grow to 0.1-0.2 cm and then the membranes are prepared for hybridization to nucleic acid sequences. Standard protocols for membrane preparation can be found in Sambrook (*Molecular Cloning, A Laboratory Manual* 2nd Ed., Cold Spring Harbor Press, Cold Spring Harbor, New York).

Selection of nucleic acid probes, or similarly PCR primers, can be done using a variety of methods, including publicly available programs such as MIT Primer 3 (Rozen and Skaletsky (1996, 1997, 1998)), for example.

Synthetic oligos or a restriction fragment containing probe sequences, from the original clone sequence, are labeled with ³³P using conventional means, such as the random hexamer priming method (High Prime kit from Boehringer Mannheim). The labeled probes are then hybridized to the arrayed filters under stringent conditions (Sambrook *et al.*, 1989, *Molecular Cloning, A Laboratory Manual* 2nd Ed., Cold Spring Harbor Press, Cold Spring Harbor, New York). In

this case, the wash conditions can approach 70-80°C at 0.1 M NaCl. After washing, the positive colonies can be visualized by exposing the blots to film or to a phosphoimaging screen. The processed image reveals colonies containing cDNA clones of the targeted gene. Alternatively, a non-radioactive method might be chosen to label the probes and detect colony hybridization (for example, the DIG Non-Radioactive DNA Labeling and Detection Kit from Boehringer Mannheim). Each clone is analyzed by restriction digests to identify the longest clone in the group. The longer clones are sequenced using an automated sequencing system (for example a Perkin-Elmer ABI 377) and the sequences evaluated for a complete open reading frame initiated with a methionine start codon (ATG). Genomic clones can also be analyzed for intron-exon boundaries using known methods as well as compared to relevant or homologous gene or protein sequence information to determine the coding regions.

Example 3

Expression and Purification of a Polypeptide in *E. coli* Host

The vector pProEX HT is used for expression of a polypeptide in a bacterial host system. (Life Technologies Inc., Gaithersburg, MD). The plasmid encodes the ampicillin resistance gene ("Apr") and contains a pBR322 origin of replication (ori), an IPTG inducible promoter, a ribosome binding site, and six codons encoding a histidine tag at the amino terminus. The his tag allows affinity purification using immobilized metal ion affinity chromatography (IMAC), such as with the nickel-nitrilo-tri-acetic acid (Ni-NTA) affinity resin. The cloning region contains suitable restriction enzyme cleavage sites for insertion of polypeptide encoding sequences. The vector also encodes a TEV (Tobacco Etch Virus) protease cleavage site to remove the his tag region from the amino terminus of the expressed polypeptide.

The desired polypeptide encoding sequence or fragment, typically lacking any hydrophobic leader sequence, is PCR amplified from a cDNA clone. Primers

-80-

designed from at least one of SEQ NO:1-82, which anneal to the amino terminal sequences of the desired polypeptide encoding region, and from the 3' end of the cDNA clone, preferably within or at the beginning of vector sequences of the cDNA clone, are used. Additional sequence containing restriction sites, to facilitate cloning into pProEX HT vector, or stop codons can be incorporated into one or both of the primers.

PCR amplified sequences and the vector are digested with appropriate restriction enzymes and then ligated together. Insertion of the DNA into the digested pProEX HT vector places a polypeptide encoding sequence downstream from the *trc* promoter and in proper reading frame to an initiator AUG.

The ligation mixture is transformed into competent *E. coli* cells using standard procedures. Sambrook, *et al.*, 1989; Ausubel, 1989, supplements through September 1998. *E. coli* strain DH5a is used, however other strains are possible. Amp resistant colonies indicate successful transformation. Plasmid DNA from resistant colonies is isolated and the correct construction confirmed by restriction analysis, PCR, and DNA sequencing.

Clones containing the desired constructs are grown over night in LB media supplemented with amp (100 mg/ml). These culture are used to inoculate production cultures, where the cells are grown at 37°C to an OD₅₉₀ density of between about 0.4 and 0.6 before induction by adding isopropyl- β -D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM, and allowed an additional 3 to 4 hours to express polypeptide. Cells then are harvested by centrifugation.

The cell pellet is then brought up in 6M guanidine-HCl, pH8, at 40°C, and stirred for 4 hours. The cell debris is removed by centrifugation and the supernatant containing the expressed polypeptide is dialyzed against a refolding buffer, such as a NaCl-based or Tris-based buffer, at about pH 6 and containing protease inhibitors. After refolding, the polypeptide is purified by IMAC and

optionally cleaved with TEV protease. The cleavage reaction is run through a size exclusion gel to purify the desired polypeptide. Purified polypeptide is stored at 4°C or frozen at -80°C. Gel electrophoresis can be used to verify production of desired polypeptide or for purification..

5

As noted above, the specific examples should not be interpreted as a limitation to the scope of the invention. Instead, they are merely exemplary embodiments one skilled in the art would understand from the entire disclosure of this invention.

10

Table 1*

<u>Seq NO</u>	<u>Hit Description</u>	<u>P Score</u>
SEQ NO 1	Human mRNA for fibronectin (FN precursor	100
SEQ NO 2	Human lumican mRNA. complete cds.	100
SEQ NO 3	Human ferritin H chain mRNA. complete cd	100
SEQ NO 4	Human osteoprotegerin (OPG) mRNA, comple	100
SEQ NO 5	Human DNA for osteopontin. complete cds.	100
SEQ NO 6	Human complement decay-accelerating fact	100
SEQ NO 7	Human SOD-2 gene for manganese superoxid	100
SEQ NO 8	Human chitinase (HUMTCHIT) mRNA, exon 1b	100
SEQ NO 9	Human CRTL1 gene exons 1, 2, 3, 4 and 5.	100
SEQ NO 10	Human glia-derived nexin (GDN) mRNA, 5'	100
SEQ NO 11	Human SPARC osteonectin mRNA, complete c	100
SEQ NO 12	Human mRNA for CD59, an LY-6-like protei	100
SEQ NO 13	Human mRNA for pro-alpha-1 type 3 collag	100
SEQ NO 14	Human mRNA for eukaryotic initiation fac	100
SEQ NO 15	Human mRNA for DIF-2 protein.	100
SEQ NO 16	Human apolipoprotein J mRNA. complete cd	100
SEQ NO 17	Human chondroitin/dermatan sulfate prote	100
SEQ NO 18	Human alpha-1-antichymotrypsin mRNA, 3'	100
SEQ NO 19	Human alpha-1-antitrypsin mRNA, complete	100
SEQ NO 20	Human mRNA for serine protease.	100
SEQ NO 21	Human mRNA for caveolin.	96
SEQ NO 22	Human GOS2 protein gene. complete cds.	100
SEQ NO 23	Human mRNA for CC chemokine LARC precurs	100
SEQ NO 24	Human mRNA for proliferation-associated	100
SEQ NO 25	Human mRNA for collagen VI alpha-2 C-ter	100
SEQ NO 26	Human fos proto-oncogene (c-fos), comple	100
SEQ NO 27	Human zinc finger transcriptional regula	100
SEQ NO 28	Human zinc finger protein (MAZ) mRNA.	100
SEQ NO 29	Human muscle creatine kinase gene (CKMM)	100
SEQ NO 30	FBLN1: fibulin-1, isoform D precursor	9
SEQ NO 31	Human skeletal muscle LIM-protein FHL1 m	100
SEQ NO 32	Human caldesmon mRNA. complete cds.	100
SEQ NO 33	Human slow skeletal muscle troponin T mR	100
SEQ NO 34	Human mRNA for nuclear envelope protein	100
SEQ NO 35	Human mRNA for titin protein (clone hh1-	100
SEQ NO 36	Human alpha-B-crystallin gene, 5' end.	100
SEQ NO 37	Human sult1s01 mRNA. complete cds.	100
SEQ NO 38	Human CAPL protein mRNA. complete cds.	100
SEQ NO 39	Human mRNA for KIAA0253 gene. partial cd	100
SEQ NO 40	Human tenascin-X (XA) mRNA, complete cds	100
SEQ NO 41	Human gene for fibromodulin.	100
SEQ NO 42	Human RASF-A PLA2 mRNA. complete cds.	100
SEQ NO 43	Human mRNA for activin beta-A, exon1,2.	100

-83-

SEQ NO 44	Human tumor necrosis factor-inducible (T	100
SEQ NO 45	nucleotide pyrophosphohydrolase	100
SEQ NO 46	Human Pig12 (PIG12) mRNA, complete cds.	100
SEQ NO 47	Human mRNA for cysteine dioxygenase, com	100
SEQ NO 48	Human sialoprotein mRNA, complete cds.	100
SEQ NO 49	Human mRNA for alpha-1 type II collagen.	100
SEQ NO 50	Human alpha-2 type XI collagen mRNA (COL	100
SEQ NO 51	F08D12.3	4
SEQ NO 52	Rat chondroadherin mRNA, complete cds.	76
SEQ NO 53	ChM-I: chondromodulin-I	76
SEQ NO 54	mimecan	95
SEQ NO 55	Human chondroadherin gene, 5'flanking re	100
SEQ NO 56	Human bone morphogenetic protein 2A (BMP	30
SEQ NO 57	Human (lambda) DNA for Ig light chain.	17
SEQ NO 58		0
SEQ NO 59		0
SEQ NO 60		0
SEQ NO 61		0
SEQ NO 62		0
SEQ NO 63		0
SEQ NO 64		0
SEQ NO 65		0
SEQ NO 66		0
SEQ NO 67		0
SEQ NO 68		0
SEQ NO 69		0
SEQ NO 70		0
SEQ NO 71		0
SEQ NO 72		0
SEQ NO 73		0
SEQ NO 74		0
SEQ NO 75		0
SEQ NO 76		0
SEQ NO 77	Human mRNA for scrapie responsive protein	78
SEQ NO 78		0
SEQ NO 79		0
SEQ NO 80		0
SEQ NO 81		0
SEQ NO 82		0

-84-

Table 2

<u>Seq NO</u>	<u>Hit Description</u>	<u>P Score</u>
SEQ NO 1	Human mRNA for fibronectin (FN precursor	100
SEQ NO 2	Human lumican mRNA, complete cds.	100
SEQ NO 3	Human ferritin H chain mRNA, complete cd	100
SEQ NO 4	Human osteoprotegerin (OPG) mRNA, comple	100
SEQ NO 5	Human DNA for osteopontin, complete cds.	100
SEQ NO 6	Human complement decay-accelerating fact	100
SEQ NO 8	Human chitinase (HUMTCHIT) mRNA, exon 1b	100
SEQ NO 9	Human CRTL1 gene exons 1, 2, 3, 4 and 5.	100
SEQ NO 10	Human glia-derived nexin (GDN) mRNA, 5'	100
SEQ NO 11	Human SPARC/osteonectin mRNA, complete c	100
SEQ NO 12	Human mRNA for CD59, an LY-6-like protei	100
SEQ NO 30	FBLN1; fibulin-1, isoform D precursor	9

We claim:

1. A nucleic acid having a nucleotide sequence selected from the group
5 consisting of SEQ NO: 1 through SEQ NO: 82, or complements thereof.
2. The nucleic acid capable of specifically hybridizing to a nucleic acid of
claim 1, or a complement thereof.
3. The nucleic acid exhibiting a percentage identity of between about 70% to
about 90% with at least a 10 nucleotide region of the sequence of a nucleic
10 acid of claim 2.
4. The nucleic acid exhibiting a percentage identity of between about 90% to
about 99% with at least a 10 nucleotide region of the sequence of a nucleic
acid of claim 3.
5. The nucleic acid as claimed in any one of claims 3 or 4 wherein said
15 nucleic acid is detectably labeled.
6. The nucleic acid of claim 5 wherein said sequence is a marker of
osteoarthritis progression.
7. The nucleic acid of claim 5 wherein said label is selected from the group
consisting of radioactive, fluorescent, chemi-luminescent, and
20 chromogenic agents, and magnetic particles.
8. A method of identifying a nucleic acid comprising contacting a
hybridization probe as claimed in claim 5 with a sample containing nucleic
acid and detecting hybridization to the hybridization probe.
9. A method of identifying a nucleic acid comprising contacting a PCR probe
25 as claimed in claim 5 with a sample containing nucleic acid and producing
multiple copies of a nucleic acid that hybridizes to the PCR probe.

10. A substantially-purified nucleic acid having at least one 10 nucleotide region substantially identical to a sequence identified in Table 1.
11. A recombinant DNA comprising a nucleic acid according to one of claims 1-4, or 8, wherein the recombinant nucleic acid further comprises a promoter or
5 partial promoter region.
12. A host cell containing a nucleic acid as claimed in claim 10.
13. A method for producing and purifying a polypeptide, said method comprising the steps of:
- a) culturing the host cell of claim 12 under conditions suitable for the
10 expression of the peptide; and
- b) recovering the polypeptide from the host cell culture.
14. A substantially-purified protein, polypeptide, or fragment thereof, wherein at least one 15 amino acid region is encoded by a nucleic acid as claimed in one of claims 1-4, or 9.
15. An antibody that specifically binds to a purified protein, polypeptide, or fragment thereof, having at least one region of 5 contiguous amino acids encoded by a nucleic acid as claimed in one of claims 1-4 or 10.
16. A transgenic animal having in one or more of its cells an introduced nucleic acid as claimed in one of claims 1-4 or 10, or progeny of the transgenic
20 animal.
17. A cell taken from a transgenic animal or its progeny as claimed in claim 16.
18. A composition comprising a nucleic acid as claimed in one of claims 1-3, or a complement thereof.
19. A method of identifying a biologically active compound or composition comprising contacting the compound or composition with a sample comprising a

protein, polypeptide, or fragment as claimed in 13, and comparing the interaction between the compound or composition and the protein, polypeptide, or fragment with a control.

20. A compound or composition that is detectable in a method of claim 18.
- 5 21. A computer-readable medium having recorded thereon the sequence information of one or more of SEQ NO:1 through SEQ NO: 82 or complements thereof.
22. A method of identifying a nucleic acid comprising providing a computer-readable medium as claimed in claim 21 and comparing nucleotide
10 sequence information using a computerized means.
23. A substantially-purified nucleic acid molecule which comprises a nucleic acid sequence that is identical to at least 10 nucleotides of a nucleotide sequence selected from the group consisting of SEQ NO: 1 through SEQ NO: 82, or complements thereof.
- 15 24. A substantially-purified nucleic acid molecule which comprises a nucleic acid sequence that is identical to at least 50 nucleotides of a nucleotide sequence selected from the group consisting of SEQ NO: 1 through SEQ NO: 82, or complements thereof.
25. A substantially-purified nucleic acid molecule which comprises a nucleic
20 acid sequence that is identical to at least 100 nucleotides of a nucleotide sequence selected from the group consisting of SEQ NO: 1 through SEQ NO: 82, or complements thereof.
26. A substantially-purified protein, polypeptide, or fragment thereof, of claim 13 wherein said substantially-purified protein, polypeptide, or fragment thereof is
25 a fusion protein.
27. An antibody of claim that is detectably-labeled.

28. A transformed cell having a nucleic acid molecule of claim 1.
29. A transformed cell having the antisense of a nucleic acid molecule of claim 1.
30. A process for diagnosis or prognosis of osteoarthritis a mammal from the
5 expression of mRNA or cDNA that is identical to at least 20 nucleotides of
a nucleotide sequence selected from the group consisting of SEQ NO: 1
through SEQ NO: 82, or complements thereof.
31. A method of isolating a nucleic acid that is identical to at least 20
nucleotides of a nucleotide sequence selected from the group consisting of
10 SEQ NO: 1 through SEQ NO: 82, or complements thereof.

<110> Searle/Monsanto
Phippard, Deborah
Vasanthakamur, Geetha
Dotson, Stanton
Ma, Xiao-Jun

<120> Osteoarthritis tissue-derived nucleic acids, polypeptides,
vectors, and cells

<130> SO-3221 PR

<160> 82

<210> 1
<211> 310
<212> DNA
<213> Homo sapiens

<400> 1

```
cagaataact ctttctgcac agaccacact gttttggttc agactcgagg aggaaattcc 60
aatggtgcct tgtgccactt ccccttccta tacaacaacc acaattacac tgattgcact 120
tctgagggca gaagagacaa catgaagtgg tgtgggacca cacagaacta tgatgccgac 180
cagaagtttg ggttctgccc catggctgcc cagaggaaa tctgcacaac caatgaaggg 240
gtcatgtacc gcattggaga tcagtgggat aagcagcatg acatgggttc acatgatgag 300
gtgcacgttt                                     310
```

<210> 2
<211> 1986
<212> DNA
<213> Homo sapiens

<400> 2

```
cttgggctgt cctttctccc cacgttcacc tgcacttcgt tagagagcag tgttcacatg 60
ccacaccaca agatccccac aatgacataa ctccattcag agactggcgt gactgggctg 120
ggctctccca ccccccttca gctcttgat cactcagaat ctggcagcca gttccgtcct 180
gacagagttc acagcatata ttggtggatt cttgtccata gtgcatctgc tttagaatt 240
aacgaaagca gtgtcaagac agtaaggatt caaaccattt gccaaaaatg agtctaagt 300
catttactct ctctctggca ttgattggtg gtaccagtgg ccagtactat gattatgatt 360
ttccctatc aatttatggg caatcatcac caaactgtgc accagaatgt aactgccctg 420
aaagctaccc aagtgccatg tactgtgatg agctgaaatt gaaaagtgt ccaatgggtg 480
ctcttggaat caagtatctt taccttagga ataaccagat tgaccatatt gatgaaagg 540
cctttgagaa tgtaactgat ctgcagtggc tcattctaga tcacaacctt ctagaaaact 600
ccaagataaa agggagagtt ttctctaaat tgaaacaact gaagaagctg catataaacc 660
acaacaacct gacagagtct gtgggcccac ttcccaaacc tctggaggat ctgcagctta 720
```

ctcataacaa gatcacaaaag ctgggctctt ttgaaggatt ggtaaacctg accttcatcc 780
 atctccagca caatcggtg aaagaggatg ctgtttcagc tgcttttaaa ggtcttaaat 840
 cactcgaata ccttgacttg agcttcaatc agatagccag actgccttct gggctctcct 900
 gtctctcttc taactctcta cttagacaac aataagatca gcaacatccc tgatgagtat 960
 ttcaagcgtt ttaatgcatt gcagtatctg cgtttatctc acaacgaact ggctgatagt 1020
 ggaataacctg gaaattcttt caatgtgtca tccctgggtg agctggatct gtcctataac 1080
 aagcttaaaa acataccaac tgtcaatgaa aaccttgaaa actattacct ggaggtcaat 1140
 caacttgaga agtttgacat aaagagcttc tgcaagatcc tggggccatt atcctactcc 1200
 aagatcaagc atttgcggtt ggatggcaat cgcattctcag aaaccagtct tccaccggat 1260
 atgtatgaat gtctacgtgt tgctaacgaa gtcactctta attaatatct gtatcctgga 1320
 acaatatctt atgggtatgt ttttctgtgt gtcagtttct atagtatcca tattttatta 1380
 ctgtttatta ctccatgaa ttttaaaatc tgagggaaat gttttgtaaa cattttattt 1440
 ttttaagaa aagatgaaag gcaggcctat ttcatacaca gaacacacac atatacacga 1500
 atagacatca aactcaatgc tttatttgta aatttagtgt ttttttattt ctactgtcaa 1560
 atgatgtgca aaacctttta ctggttgcat ggaaatcagc caagttttat aatccttaaa 1620
 tcttaatggt cctcaaagct tggattaaat acatatggat gttactctct tgcaccaa 1680
 tatcttgata cattcaaatt tgtctgggtt aaaaataggt ggtagatatt gaggccaaga 1740
 atattgcaaa atacatgaag cttcatgcac ttaaagaagt atttttagaa taagaatttg 1800
 catacttacc tagtgaaact tttctagaat tatttttcac tctaagtcac gtatgtttct 1860
 ctttgattat ttgcatgtta tgtttaataa gctactagca aaataaaaca tagcaaatgg 1920
 catcactgtg ttgacttct tgtgaaattt ctgtactttg tatataaaat acataaaaca 1980
 atagat 1986

<210> 3
 <211> 920
 <212> DNA
 <213> Homo sapiens
 <400> 3

ccgagagtcg tcgggggttc ctgcttcaac agtgcttggg cggaaccggg cgctcgttcc 60
 ccaccccggc cgcccgccca tagccagccc tccgtcacct ctccaccgca ccctcggact 120
 gccccaaggc ccccgccgccc gctccagcgc cgcgcagcca ccgcccgcgc cgccgcctct 180
 ccttagtcgc cgccatgacg accgcgtcca cctcgcaggt gcgccagaac taccaccagg 240
 actcagaggc cgccatcaac cgccagatca acctggagct ctacgcctcc tacgtttacc 300
 tgtccatgtc ttactacttt gaccgcgatg atgtggcttt gaagaacttt gccaaatact 360

ttcttcacca atctcatgag gagaggggaac atgctgagaa actgatgaag ctgcagaacc 420
 aacgaggtgg ccgaatcttc cttcaggata tcaagaaacc agactgtgat gactgggaga 480
 gcggtctgaa tgcaatggag tgtgcattac atttggaaaa aaatgtgaat cagtcactac 540
 tggaactgca caaactggcc actgacaaaa atgaccccca tttgtgtgac ttcattgaga 600
 cacattacct gaatgagcag gtgaaagcca tcaaagaatt ggggtgaccac gtgaccaact 660
 tgcgcaagat gggagcgccc gaatctggct tggcggaata tctctttgac aagcacaccc 720
 tgggagacag tgataatgaa agctaagcct cgggctaatt tccccatagc cgtggggtga 780
 cttccctggt caccaaggca gtgcatgcat gttggggttt cctttacctt ttctataagt 840
 tgtacaaaa catccactta agttctttga tttgtaccat tccttcaaat aaagaaattt 900
 ggtacccagg aaaaaaaaaa 920

<210> 4
 <211> 2139
 <212> DNA
 <213> Homo sapiens

<400> 4
 caggcgatac ttcctgttgc cgggacgcta tatataacgt gatgagcgca cgggctgcgg 60
 agacgcaccg gagcgctcgc ccagccgccc cctccaagcc cctgagggtt ccggggacca 120
 caatgaacaa cttgctgtgc tgcgcgcttc gtgtttcttg acatctccat taagtggacc 180
 acccagga aa cgtttcctcc aaagtacctt cattatgacg aagaaacctc tcatcagctg 240
 ttgtgtgaca aatgtcctcc tgggtacctac ctaaaacaac actgtacagc aaagtgggaa 300
 accgtgtgcg ccccttgcgc tgaccactac tacacagaca gctggcacac cagtgcagag 360
 tgtctatact gcagcccgt gtgcaaggag ctgcagtacg tcaagcagga gtgcaatcgc 420
 acccacaacc gcgtgtgcga atgcaaggaa gggcgctacc ttgagataga gttctgcttg 480
 aaacatagga gctgccctcc tggatttggg gtggtgcaag ctggaacccc agagcgaat 540
 acagtttgca aaagatgtcc agatgggttc ttctcaaag agacgtcatc taaagcacc 600
 tgtagaaaac acacaaattg cagtgtcttt ggtctcctgc taactcagaa aggaaatgca 660
 acacacgaca acatatgttc cggaaacagt gaatcaactc aaaaatgtgg aatagatgtt 720
 accctgtgtg aggaggcatt cttcagggtt gctgttecta caaagtttac gcctaactgg 780
 cttagtgtct tggtagacaa tttgcctggc accaaagtaa acgcagagag ttagagagg 840
 ataaaacggc aacacagctc acaagaacag actttccagc tgctgaagtt atggaaacat 900
 caaaacaaag accaagatat agtcaagaag atcatccaag atattgacct ctgtgaaaac 960
 agcgtgcagc ggcacattgg acatgctaac ctcaccttcg agcagcttcg tagcttgatg 1020
 gaaagcttac cgggaaagaa agtgggagca gaagacattg aaaaaacaat aaaggcatgc 1080

aaaccagtg accagatcct gaagctgctc agtttgtggc gaataaaaaa tggcgaccaa 1140
 gacacctga agggcctaata gcacgcacta aagcactgca aagacgtacc actttcccaa 1200
 aactgtcact cagagtctaa agaagacat cagggttcctt cacagcttca caatgtacaa 1260
 attgtatcag aagttatttt tagaaatgat aggtaaccag gtccaatcag taaaaataag 1320
 ctgcttataa ctggaaatgg ccattgagct gtttctctac aattggcgag atcccatgga 1380
 tgagtaaaact gtttctcagg cacttgaggc tttcagtgat atctttctca ttaccagtga 1440
 ctaattttgc cacagggtac taaaagaaac tatgatgtgg agaaaggact aacatctcct 1500
 ccaataaacc ccaaatgggt aatccaactg tcagatctgg atcgttatct actgactata 1560
 ttttccctta ttactgcttg cagtaattca actggaaatt aaaaaaaaaa aactagactc 1620
 cattgtgctt tactaaatat gggaaatgtct aacttaaata gctttgagat ttcagctatg 1680
 ctaggaggctt ttattagaaa gccatatttt tttctgtaaa agttactaat atatctgtaa 1740
 cactattaca gtattgctat ttatattcat tcagatataa gatttgtaca tattatcatt 1800
 ctataaagaa acggtatgac ttaatttttag aaagaaaatt atattctgtt tattatgaca 1860
 aatgaaagag aaaatatata tttttaatgg aaagtttga gcatttttct aatagggtact 1920
 gccatatttt tctgtgtgga gtatttttat aattttatct gtataagctg taatatcatt 1980
 ttatagaaaa tgcattattt agtcaattgt ttaatgttgg aaaacatag aaatataaat 2040
 tatctgaata ttagatgctc tgagaaattg aatgtacctt atttaaaaga ttttatgggt 2100
 ttataactat ataaatgaca ttattaaagt tttcaaatt 2139

<210> 5
 <211> 157
 <212> DNA
 <213> Homo sapiens

<400> 5
 cccaatacta agctcctctg gttagagcca gccatgagag aaactccaag tacttctgac 60
 tggttctctc tctactcacc cacccttag gtggtgcag aaggaaactct gtgcaacccc 120
 cagagttctc attctcagtg acagggaaat gtaatga 157

<210> 6
 <211> 2263
 <212> DNA
 <213> Homo sapiens
 <223> unsure at all n locations
 <400> 6

acctctgacc acaacaaacc cctactccac ccggtcttgt ttgtcccacc cttgggtgacg 60
 cagagcccca gccagacccc cgcccaaagc actcatttaa ctgggtattgc ggancacgag 120

gcttctgctt actgcaactc gctccggccg ctgggcgtag tgcgactcgg cggagtcgcc 180
gcgggcgcgtc cttgttctaa cccggcgcg ccatgaccgtc gcgcggccga gcgtgccgc 240
ggcgctgccc ctccctgggg agctgccccg gctgctgctg ctgggtgctgt tgtgcctgcc 300
ggccgtgtgg ggtgactgtg gccttcccc agatgtacct aatgcccagc cagctttgga 360
aggccgtaca agttttcccc aggatactgt aataacgtac aaatgtgaag aaagctttgt 420
gaaaattcct ggcgagaagg actcagtgat ctgccttaag ggcagtcaat ggtcagatat 480
tgaagagttc tgcaatcgta gctgcgaggt gccacaagg ctaaattctg catccctcaa 540
acagccttat atcactcaga attattttcc agtcgggtact gttgtggaat atgagtgccg 600
tccaggttac agaagagaac cttctctatc accaaaacta acttgccctc agaatttaa 660
atgggtccaca gcagtcgaat tttgtaaaaa gaaatcatgc cctaataccgg gagaaatacg 720
aaatggtcag attgatgtac cagggtggcat attatttgg gcaaccatgc tccttctcat 780
gtaacacagg gtacaaatta tttggctcga cttctagttt ttgtcttatt tcaggcagct 840
ctgtccagtg gagtgaccgg ttgccagagt gcagagaaat ttattgtcca gcaccaccac 900
aaattgacaa tggaataatt caaggggaac gtgaccatta tggatataga cagtctgtaa 960
cgtatgcatg taataaagga ttcaccatga ttggagagca ctctatttat tgtactgtga 1020
ataatgatga aggagagtgg agtggcccac cacctgaatg cagaggaaaa tctctaactt 1080
ccaagggtccc accaacagtt cagaaaccta ccacagtaaa tgttccaact acagaagtct 1140
caccaacttc tcagaaaacc accacaaaaa ccaccacacc aaatgctcaa gcaacacgga 1200
gtacacctgt ttccaggaca accaagcatt ttcatgaaac aaccccaaat aaaggaagt 1260
gaaccacttc aggtactacc cgtcttctat ctgggcacac gtgtttcacg ttgacagggt 1320
tgcttgggac gctagtaacc atgggcttgc tgacttagcc aaagaagagt taagaagaaa 1380
atacacacaa gtatacagac tgttcctagt ttcttagact tatctgcata ttggataaaa 1440
taaattgcaat tgtgctcttc atttaggatg ctttcattgt ctttaagatg tgttaggaat 1500
gtcaacagag caaggagaaa aaaggcagtc ctggaatcac attcttagca cacctacacc 1560
tcttgaaaat agaacaactt gcagaattga gagtgattcc tttcctaaaa gtgtaagaaa 1620
gcatagagat ttgttcgtat ttagaatggg atcacgagga aaagagaagg aaagtgattt 1680
ttttccacaa gatctgtaat gttatttcca cttataaagg aaataaaaaa tgaaaaacat 1740
tatttggata tcaaaagcaa ataaaaacc aattcagctc cttctaagca aaattgctaa 1800
agagagatga accacattat aaagtaatct ttggctgtaa ggcattttca tctttccttc 1860
gggttggcaa aatattttta aggtaaaaca tgctggtgaa ccaggggtgt tgatggtgat 1920
aagggaggaa tatagaatga aagactgaat cttcctttgt tgcacaaata gagtttggaa 1980
aaagcctgtg aaaggtgtct tctttgactt aatgtcttta aaagtatcca gagatactac 2040

6

aatattaaca taagaaaaga ttatatatta tttctgaatc gagatgtcca tagtcaaatt 2100
 tgtaaatctt attcttttgt aatatttatt tatatttatt tatgacagt aacattctga 2160
 ttttacatgt aaaacaagaa aagttgaaga agatatgtga agaaaaatgt atttttccta 2220
 aatagaaata aatgatccca ttttttgga aaaaaaaaaaaa aaa 2263

<210> 7
 <211> 712
 <212> DNA
 <213> Homo sapiens

<400> 7

cttaaaccta ttttagtaatg ttttcccaag tttattttt atttttaatt ttttcccaa 60
 gtttattttt ctattttttt ttcattgaaa aatggggtaa cttagcagtt tcaatattga 120
 agactgaagt ttaaaaaaaaa ttttaattca aggtactttt aaaattcagt tagaaaagta 180
 ggctttaaaa attattagag acaagagtac caaagcggg tgggtatgtg tgtgtgtgta 240
 tgcattgctg tggattggaa aaactttgga gactgattac ttttcattat atatgtgtca 300
 cagtgaaca gcttttatgt gtcattgaag attattgctt gcctctctaa ggaaggctcg 360
 gactgtttta atagacgggc aaggtggaac cttttgaaag atgagctttt gaataaagt 420
 tgtctgctag atcattggtt gtattgaact aacaagggtt gcagatctgc tgacttatat 480
 aaagcttttt gattcctact aagctttaag atttaaaaaa tgttcaatgt tgaaatttct 540
 gtggggctct atttttgctt tggctttctg gtgagagagt gaggaagcat tctttccttc 600
 actaagtttg tctttcttgt cttctggata gattgattt aagagactaa gggaatttac 660
 aaactaaaga ttttagtcat ctggtggaag agggagactt aagattgttt ag 712

<210> 8
 <211> 1474
 <212> DNA
 <213> Homo sapiens

<400> 8

ctcagtggat aaaagacctg gagaatgtg atccagaag aagctggcca aggatatggg 60
 agcaaccacc atgggaccag aagctctctt ggggcagggt tagtggctctt gctgcttctc 120
 caggagggga tctgcctaca aactgggttg cttctttacc aactgggtcc caggaccggc 180
 aggaaccagg aaaattcacc cctgaggaat attgacctt tcttatgtc tcatctcacc 240
 tattcattgc gccagcatg aaaacaacaa ggttatcacc aaggacaaga gtgaagtgat 300
 gctctaccag accatcaaca gttctcaaaa ccaagaatcc caaactgaaa attctcttct 360
 ccattggagg gtactgtttt ggttccaaag ggttccacct tatgggtgat tcttctacat 420
 cactgttga attcattaac tccataatcc tgtttctgag gaaccataac tttgatggac 480

7

tggatgtaag ctggatctac ccagatcaga aagaaaacac tcatttcact gtgctgattc 540
 atgagtttagc agaagccttt cagaaggact tcacaaaatc caccaaggaa aggcttctct 600
 tgactgcggg gggatatctgc agggaggcaa atgattgata acagctatca agttgagaaa 660
 ctggcaaaaag atctggattt catcaacctc ctgtcctttg acttccatgg gtcttgggaa 720
 aagccctta tcaactggcca caacagccct gctgagcaag ggggtggcagg acagagggcc 780
 aagctcctac tacaatgtgg aatatgctgt ggggtactgg atacataagg gaatgccatc 840
 agagaagggtg gtcatgggca tccccacata tggggcactc cttcacactg gcctctgcag 900
 aaaccaccgt gggggccctt gcctctggcc ctggagctgc tggaccatc acagagtctt 960
 caggcttctt ggccatttat gagatctgcc agttcctgaa aggagccaag atcacgcggc 1020
 tccaggatca gcaggttccc tacgcagtca aggggaacca gtgggtgggc tatgatgatg 1080
 tgaagagtat ggagaccaag gttcagttct taaagaattt aaacctggga ggagccatga 1140
 tctggtctat tgacatggat gacttcaactg gcaaactctg caaccagggc ccttaccctc 1200
 ttgtccaagc agtcaagaga agccttggct cccgtgaag gattaactta cagagaagca 1260
 ggcaagatga ccttgcctgc tggggcctgc tctctcccag gaattctcat gtgggattcc 1320
 ccttgcagg ccggcctttg gatctctctt ccaagccttt cctgacttcc tcttagatca 1380
 tagattggac ctgggtttgt tttcctgcag ctgttgactt gttgccctga agtacaataa 1440
 aaaaaattca ttttgctcca gtaaaaaaaa aaaa 1474

<210> 9
 <211> 592
 <212> DNA
 <213> Homo sapiens
 <223> unsure at all n locations
 <400> 9

actttcctgg tgacgctttg cttttcttct gctcttgggtg agaaagtgcc tccttcttcc 60
 caggatcagg acctctgcca tccagcgcca caaagagaca tttctgcaca cacactnnnn 120
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nncagagac aaacttaagg tgaggagaaa 180
 gagcgctagt ttcaattgat ctccagcttc caacttaagc agaacttgag agcatccgaa 240
 ctctctggatt tcaggacaag tgaagaagat tctttgggct ataagatga agagtctact 300
 tcttctgggtg ctgatttcaa tctgctgggc tgatcatctt tcagacaact atactctgga 360
 tcatgacaga gctattcaca tccaagcaga aaatgggccc ccatctactt gtggaagcag 420
 agcaagccaa ggtgttttca caccagagggt ggcaatgtta cactgccatg taaattttat 480
 cgagacccta cagcattttg ctccaggaatc cataaaatcc gaattaagtg gaccaagcta 540
 acttcggatt acctcaagga agtggatgtt tttgtttcca tgggatacca ca 592

<210> 10
<211> 2004
<212> DNA
<213> Homo sapiens

<400> 10

gcgaccgccc cctgtgatcc agcgagcgcg gtcgtccttg gtggaaggaa ccatgaactg 60
gcattctccc ctcttctctt tggcctctgt gacgctgcct tccatctgct ccactttcaa 120
tcctctgtct ctcgaggaa taggctccaa cacggggatc caggttttca atcagattgt 180
gaagtcgagg cctcatgaca acatcgtgat ctctcccat gggattgctg cggtcctggg 240
gatgcttcag ctgggggagg acggcaggac caagaagcag ctcgccatgg tgatgagata 300
cggcgtaaat ggagttggtg aaatattaaa gaagatcaac aaggccatcg tctccaagaa 360
gaataaagac attgtgacag tggctaacgc cgtgtttgtt aagaatgcct ctgaaattga 420
agtgcctttt gttacaagga acaagatgt gttccagtgt gaggtccgga atgtgaactt 480
tgaggatccc agcctctgcc tgtgattcca tcaatgcatg ggttaaaaac gaaaccaggg 540
atatgattga caatctgctg tcccagatc ttattgatgg tgtgctcacc agactggtec 600
tcgtcaacgc agtgattttc aagggtctgt ggaaatcacg gttccaaccc gagaacacaa 660
agaaacgcac ttctgtggca gccgacggga aatcctatca agtgccaatg ctggcccagc 720
tctccgtgtt ccggtgtggg tcgacaagt ccccaatga tttatggtac aacttcattg 780
aactgccta ccacggggaa agcatcagca tgetgattgc actgccgact gagagctcca 840
ctcgcgtgtc tgccatcatc ccacacatca gcaccaagac catagacagc tggatgagca 900
tcatgggtgc caagaggggtg cagggtgatc tgcccaagtt cacagctgta gcacaaacag 960
atttgaagga gccgctgaaa gttcttggca ttactgacat gtttgattca tcaaaggcaa 1020
attttgcaaa aataacaagg tcagaaaacc tccatgtttc tcatatcttg caaaaagcaa 1080
aaattgaagt cagtgaagat ggaaccaaag cttcagcagc aacaactgca attctcattg 1140
caagatcatc gcctccctgg tttatagtag acagaccttt tctgtttttc atccgacata 1200
atcctacagg tgctgtgtta ttcattggggc agataaaca accctgaaga gtatacaaaa 1260
gaaacatgc aaagcaacga ctactttgct acgaagaaag actcctttcc tgcattttc 1320
atagttctgt taaatatatt tgtacatcgc tcttttttca aaactagttc ttaggaacag 1380
actcgatgca agtgtttctg ttctgggagg tattggaggg aaaaaacaag caggatggct 1440
ggaacactgt actgaggaat gaatagaaag gcttcagat gtctaaaaga ttctttaaac 1500
tactgaactg ttacctaggc taacaaccct gttgagtatt tgctgtttgt ccagttcagg 1560
aatttttgtt ttgttttgc tatatgtgcg gcttttcaga agaaatttaa tcagtgtgac 1620
agaaaaaaaa atgttttatg gtagctttta ctttttatga aaaaaaatt atttgccttt 1680

9

taaattcttt tccccatcc cctccaaag tcttgatagc aagcggtatt ttgggggtag 1740
 aaacggtgaa atctctagcc tctttgtgtt tttgtgttg ttgtgttgt tgtttatat 1800
 aatgcatgta ttactaaaa taaaatttaa aaaactctg tcttgctaga caaggttgct 1860
 gttgtgcagt gtgcctgtca ctactggtct gtactcctg gatttgcat tttgtatttt 1920
 gtacaaagta aaaataaact gttatgagta gtaaaaata agctatttct ctgctatttg 1980
 aaaataaaaa aaaaaaaaaa aaaa 2004

<210> 11
 <211> 2128
 <212> DNA
 <213> Homo sapiens

<400> 11
 agactgccgg agagcgcgct ctgcctgcc cctgcctgcc tgccactgag gggtcccagc 60
 accatgaggg cctggatctt ctttctcctt tgccctggccg ggagggcctt ggcagcccct 120
 cagcaagaag ccctgcctga tgagacagag gtggtggaag aaactgtggc agaggtgact 180
 gaggtatctg tgggagctaa tctgtccag gtggaagtag gagaatttga tgatggtgca 240
 gaggaaccg aagaggaggt ggtggcggaa aatccctgcc agaaccacca ctgcaaacac 300
 ggcaagggtg gcgagctgga tgagaacaac acccccatgt gcgtgtgcca ggacccacc 360
 agctgccag ccccatctgg cgagtttgag aagggtgtgca gcaatgacaa caagaccttc 420
 gactcttctt gccacttctt tgccacaaag tgcaccctgg agggcaccaa gaaggccac 480
 aagctccacc tggactacat cgggccttgc aaatacatcc ccccttgctt ggactctgag 540
 ctgaccgaat tccccctgcg catgcgggac tggctcaaga acgtcctggt caccctgtat 600
 gagagggatg aggacaacaa cttctgact gagaagcaga agctgcgggt gaagaagatc 660
 catgagaatg agaagcgctt ggaggcaggg agaccacccc gtggagctgc tggcccggga 720
 cttcgagaag aactataaca tgtacatctt ccctgtacac tggcagttcg gccagctgga 780
 ccagcacccc attgacgggt acctctccca caccgagctg gctccactgc gtgtccccct 840
 catccccatg gagcattgca ccaccgctt tttcgagacc tgtgacctgg acaatgacaa 900
 gtacatcgcc ctggatgagt gggccggtg cttcggcatc aagcagaagg atatcgacaa 960
 ggatcttggt atctaaatcc actccttcca cagtaccgga ttctctcttt aacctcccc 1020
 ttcgtgtttc cccaatggt taaaatgttt ggtggtttg ttgttctgcc tggagacaag 1080
 gtgctaaca agatttaagt gaatacatta acggtgctaa aaatgaaaat tctaaccacaa 1140
 gacatgacat tcttagctgt aacttaacta ttaaggcctt ttccacacgc attaatagtc 1200
 ccatttttct cttgccattt gtagctttgc ccattgtctt attggcacat ggggtggacac 1260
 ggatctgctg ggctctgctt taaacacaca ttgcagcttc aacttttctc tttagtgttc 1320

10

tgtttgaaac taatacttac cgagtcagac tttgtgttca tttcatttca gggctcttggc 1380
 tgcctgtggg ctttccccag ggtggcctgg gaggtgggca aagggaaagta acagacacac 1440
 gatgtgttca aggatgggtt tgggactaga ggctcagtg tgggagagat ccctgcagaa 1500
 cccaccaacc agaacgtggg ttgcctgagg ctgtaactga gagaaagatt ctggggctgt 1560
 cttatgaaaa tatagacatt ctcacataag ccagttcat caccatttcc tcctttacct 1620
 ttcagtgcag tttcttttca cattaggctg ttggttcaaa cttttgggag cacggactgt 1680
 cagttctctg ggaagtggc agcgcatcct gcagggcttc tcctcctctg tcttttggag 1740
 aaccagggtt cttctcaggg gctctagggg ctgccaggct gtttcagcca ggaaggccaa 1800
 aatcaagagt gagatgtaga aagttgtaaa atagaaaaag tggagttggg gaatcggttg 1860
 ttctttcttc acatttggat gattgtcata aggttttttag catgttcctc cttttcttca 1920
 ccctcccttc tgttcttcta ttaatcaaga gaaacttcaa agttaatggg atggtcggat 1980
 ctcacaggct gagaactcgt tcacctccaa gcatttcatg aaaaagctgc ttcttattaa 2040
 tcatacaaac tctcaccatg atgtgaagag tttcacaaat ctttcaaaat aaaaagtaat 2100
 gacttagaaa ctgcaaaaaa aaaaaaaa 2128

<210> 12
 <211> 2073
 <212> DNA
 <213> Homo sapiens

<400> 12
 agtacacact ggggcttata gggactgagc ctactcaagg gtatatgggtg ctgtgggtca 60
 gagctggggc atggcaggcg attcagtgtg ccttgactcc ccctgtaaat gttcctctca 120
 gaagccttct tggccttcca gcccttgggt tttgagacaa ccagcagtc tttgttcgtt 180
 cctgacattc ctctctgtcc cttccttcca ggttctgtgg acaatcacia tgggaatcca 240
 aggaggggtc gtctgttcg ggctgtctgt cgtcctgggt gtcttctgcc attcagggtca 300
 tagcctgcag tgctacaact gtcttaaccc aactgctgac tgcaaaacag ccgtcaattg 360
 ttcactctgat tttgatgcgt gtctcattac caaagctggg ttacaagtgt ataacaagtg 420
 ttggaagttt gagcattgca atttcaacga cgtcacaacc ccgcttgagg gaaaatgagc 480
 taacgtacta ctgctgcaag aaggacctgt gtaactttaa cgaacagctt gaaaatgggtg 540
 ggacatcctt atcagagaaa acagttcttc tgctgggtgac tccatttctg gcagcagcct 600
 ggagccttca tccctaagtc aacaccagga gagcttctcc caaactcccc gttcctgcgt 660
 agtccgcttt ctcttgctgc cacattctaa agccttgata ttttccaaat ggatcctgtt 720
 gggaaagaat aaaattagct tgagcaacct ggctaagata ggggggctct gggagacttt 780
 gaagaccagt cctgtttgca ggggaagcccc acttgaagga agaagtctaa gagtgaagta 840

11

ggtgtgactt gaactagatt gcatgcttcc tcttttgctc ttgggaagac cagctttgcc 900
 agtgacagct tgagtgggtt ctctgcagcc ctcagattat ttttctctg gctccttgga 960
 tgtagtcaat tagcatcatt agtacatctt tggaggggtg ggcaggagta tatgagcatc 1020
 ctctctcaca tggaaagctt tcataaactt cagggatccc gtgttgccat ggaggcatgc 1080
 caaatgttcc atatgtgggt gtcagtcagg gacaacaaga tccttaatgc agagctagag 1140
 gacttctggc agggaagtgg ggaagtgttc cagatagcag ggcatgaaaa cttagagagg 1200
 tacaagtggc tgaataatga gtttttctc tgtctttaa ttttatatgg gctttgttat 1260
 cttccactgg aaaagtgtaa tagcatacat caatggtgtg ttaaagctat ttccttgcc 1320
 tttttttatt ggaatggtag gatattcttg ctttgccaca cacagttaca gaggtaaac 1380
 tctactacat gtgactggca gtattaagtg tgcttatttt aaatgttact ggtagaaagg 1440
 cagttcaggt atgtgtgtat atagtatgaa tgcagtgggg acaccctttg tggttacagt 1500
 ttgagacttc caaaggtcat ccttaataac aacagatctg caggggtatg ttttaccatc 1560
 tgcacccagc ctctgctaa ctctagctg actcagcata gattgtataa aatacctttg 1620
 taacggctct tagcacactc acagatgttt gaggctttca gaagctcttc taaaaaatga 1680
 tacacacctt tcacaagggc aaactttttc cttttccctg tgtattctag tgaatgaatc 1740
 tcaagattca gtagacctaa tgacatttgt attttatgat cttggctgta tttaatggca 1800
 taggctgact ttgcagatg gaggaatttc ttgattaatg ttgaaaaaaa acccttgatt 1860
 atactctgtt ggacaaaccg agtgcaatga atgatgcttt tctgaaaatg aaatataaca 1920
 agtgggtgaa tgtgggtatg gccgaaaagg atatgcagta tgcttaatgg tagcaactga 1980
 aagaagacat cctgagcagt gccagctttc ttctgttgat gccgttccct gaacatagga 2040
 aaatagaaac ttgcttatca aaacttaaaa aaa 2073

<210> 13
 <211> 253
 <212> DNA
 <213> Homo sapiens

<400> 13
 gctggctact tctcgctctg cttcatccca ctattatttt ggcacaacag gaagctgttg 60
 aaggaggatg ttcccatctt ggtcagtcct atgcggatag agatgtctgg aagccagaac 120
 catgccaaat atgtgtctgt gactcaggat ccgttctctg cgatgacata atatgtgacg 180
 atcaagaatt agactgcccc aaccagaaa ttccatttgg agaattgtgt gcagtttgcc 240
 cacagcctcc aag 253

<210> 14
 <211> 1749
 <212> DNA

12

<213> Homo sapiens

<223> unsure at all n locations

<400> 14

```

tcattgtctgc gagccaggat tcccgatcca gagacaatgg ccccgatggg atggagcccg 60
aaggcgtcca tcgagagtaa ctggaatgag attgttgaca gctttgatga catgaacctc 120
tcggagtccc ttctccgtgg catctacgcc tatggctttg agaagccctc tgccatccag 180
cagcgagcca ttctaccttg tatcaagggt tatgatgtga ttgctcaagc ccaatctggg 240
actgggaaaa cgccacatt tgccatatcg attctgcagc agattgaatt agatctaaaa 300
gccaccagg ccttggctct agcaccact cgagaattgg ctgagcagat acagaagggtg 360
gtcatggcac taggagacta catggggccc tctgttcagc cctgtatcgg gggcaccaac 420
gtgctgtctg aggtgcagaa actgcagatg gaagctcccc acatcatcgt gggtagccct 480
ggcgcgtgt ttgatatgct taaccggaga taccgtccc ccaatacat caagatgttt 540
gtactggatg aagctgacga aatgttaagc cgtggattca aggaccagat ctatgacata 600
ttccaaaagc tcaacagcaa caccaggta gtttgcgtc cagccacaat gccttctgat 660
gtgcttgagg tgaccaagaa gttcatgagg gacccattc ggattcttgt caagaaggaa 720
gagttgacct tggagggtat ccgccagttc tatatcaacg tggaaagaga ggagtggaa 780
ctggacacac tatgtgactt gtatgaaacc ctgaccatca cccaggcagt catcttcac 840
aacaccgga ggaagggtga ctggctcacc gagaagatgc atgctcgaga ttctactgta 900
tccgccatgc atggagatat ggacaaaag gaacgagacg tgattatgag ggagtctcgt 960
tctggctcta gcagagtttt gattaccact gatctgctgg ccagaggcat tgatgtgcag 1020
caggtttctt tagtcatcaa ctatgacctt ccaccaaca gggaaaacta tatccacaga 1080
atcggctcag gtggacggtt tggccgtaa ggtgtggcta ttaacatggt gacagaagaa 1140
gacaagagga ctcttcgaga cattgagacc tcttacaaca cctccattga ggaaatgcc 1200
ctcaatgttg ctgacctcat ctgaggggtt gctctgcca ccagccccag ccagggtca 1260
atctctgggg gctgaggagc agcaggagg ggaagggaag ggagccaagg gatggacatc 1320
ttgtcatttt ttttctttga ataatgtca ctttttgagg caaaagaagg aaccgtgaac 1380
attttagaca cctttttctt tggggtaggc ttttgcccc ggcnccggt cttctccnaa 1440
aaaaaaaa cactaatcca tttccctaac ctagtaacct ccagatccca gaggtctctc 1500
tcacctcagc tgagctcctt tgaagtgat tcaagggact atgtcactca gcctcatttg 1560
ctggacaaaa tctggaggga gaacctta aaccttaag tgaggttgcc cagggggttg 1620
tcccagggtg gggggaagca ggggagagaa aatggtagc atttttacat tgtttgtat 1680
agtatttatt gattcaggaa acaaacaca aattctgaat aaaatgactt ggaaactgaa 1740
aaaaaaaa

```

1749

<210> 15
 <211> 1232
 <212> DNA
 <213> Homo sapiens

<400> 15
 ttacactccg ctcggtcac catgtgtcac tctcgagct gccacccgac catgaccatc 60
 ctgcaggccc cgaccccggc cccctccacc atcccgggac cccggcggg ctccggtcct 120
 gagatcttca ccttcgaccc tctcccgag cccgcagcgg cccctgcgg gcgccccagc 180
 gcctctcgcg ggcaccgaaa gcgcagccgc agggttctct accctcgagt ggtccggcgc 240
 cagctgccag tcgaggaacc gaaccagcc aaaaggcttc tctttctgct gctcaccatc 300
 gtcttctgcc agatcctgat ggctgaagag ggtgtgccg cggccctgcc tccaagagga 360
 cgcccctaac gccgcacccc tgggcgccc cccctgtgtc ccccgctcct gagcccttta 420
 atctgacttc ggagccctcg gactacgtc tggacctcag cactttcctc cagcaacacc 480
 cggcgcgctt ctaactgtga ctcccgcac tccccaaaa gaatccgaaa aaccacaaag 540
 aaacaccagg cgtacctggt gcgcgagagc gtatcccaa ctgggacttc cgaggcaact 600
 tgaactcaga aactacagc ggagacgcca cccggtgctt gaggcgggac cgaggcgac 660
 agagaccgag gcgcatagag accgaggcac agcccagctg ggggctaggc ccggtgggaa 720
 ggagagcgtc gttaatttat ttcttattgc tcctaattaa tatttatatg tatttatgta 780
 cgtctccta ggtgatggag atgtgtacgt aatatttatt ttaacttatg caaggggtgtg 840
 agatgttccc cctgctgtaa atgcaggctc cttgggtattt attgagcttt gtgggactgg 900
 tggaaagcagg acacctggaa ctgcggcaaa gtaggagaag aaatggggag gactcgggtg 960
 ggggaggacg tcccggctgg gatgaagtct ggtgggtggg cgtaagttaa ggaggtgact 1020
 gcctcctcca gcctctcaac tccgtctgtc tactgtgtga gacttcggcg gaccattagg 1080
 aatgagatcc gtgagatcct tccatcttct tgaagtcgcc tttaggggtg ctgcgaggtta 1140
 gaggggtggg ggttgggtgg ctgtcacgga gcgactgtcg agatcgcta gtatgttctg 1200
 tgaacacaaa taaaattgat ttactgtctg ca 1232

<210> 16
 <211> 1678
 <212> DNA
 <213> Homo sapiens

<400> 16
 gtcgccagga ggagcgcgc ggcacagggt gcgctgaccg aggcgtgcaa agactccaga 60
 attggaggca tgatgaagac tctgtgtctg tttgtggggc tgctgtgtac ctgggagagt 120
 gggcagggtc tgggggacca gacggtctca gacaatgagc tccaggaaat gtccaatcag 180

ggaagtaagt acgtcaataa ggaaattcaa aatgctgtca acgggggtgaa acagataaag 240
 actctcatag aaaaaacaaa cgaagagcgc aagacactgc tcagcaacct agaagaagcc 300
 aagaagaaga aagaggatgc cctaaatgag accaggggaat cagagacaaa gctgaaggag 360
 ctcccaggag tgtgcaatga gaccatgatg gccctctggg aagagtgtaa gccctgcctg 420
 aaacagacct gcatgaagtt ctacgcacgc gtctgcagaa gtggctcagg cctggttggc 480
 cgccagcttg aggagttcct gaaccagagc tcgcccttct acttctggat gaatggtgac 540
 cgcatcgact ccctgctgga gaacgaccgg cagcagacgc acatgctgga tgtcatgcag 600
 gaccacttca gccgcgcgtc cagcatcata gacgagctct tccaggacag gttcttcacc 660
 cgggagcccc aggataccta ccactacctg cccttcagcc tgccccaccg gaggcctcac 720
 ttcttcttctt ccaagtcccg catcgtccgc agctttgatg cccttctctc cgtacgagcc 780
 cctgaacttc caccgcatgt tccagccctt ccttgagatg atacacgagg ctacgcaggc 840
 catggacatc cacttccata gcccgccctt ccagcaccgg ccaacagaat tcatacgaga 900
 aggcgacgat gaccggactg tgtgccggga gatccgccac aactccacgg gctgctgcg 960
 gatgaaggac cagtgtgaca agtgccggga gatcttgtct gtgggactgt tccaccaaca 1020
 accctcccca ggctaagctg cggcgggagc tcgacgaatc cctccaggtc gctgagaggt 1080
 tgaccaggaa atacaacgag ctgctaaagt cctaccagtg gaagatgctc aacacctcct 1140
 ccttgctgga gcagctgaac gagcagttta actgggtgtc ccggctggca aacctcacgc 1200
 aaggcgaaga ccagtactat ctgcgggtca ccacgggtggc ttccacact tctgactcgg 1260
 acgttccttc cgggtgcact gaggtggctg tgaagctctt tgactctgat cccatcactg 1320
 tgacgggtccc tgtagaagtc tccaggaaga accctaaatt tatggagacc gtggcggaga 1380
 aagcgctgca ggaataccgc aaaaagcacc gggaggagtg agatgtggat gttgcttttg 1440
 cacctacggg ggcactctgag tccagctccc cccaagatga gctgcagccc cccagagaga 1500
 gctctgcacg tcaccaagta accaggcccc agcctccagg cccccaactc cgccagcct 1560
 ctccccgctc tggatcctgc actctaacac tcgactctgc tgctcatggg aagaacagaa 1620
 ttgctcctgc atgcaactaa ttcaataaaa ctgtcttgtg agctgaaaaa aaaaaaaa 1678

<210> 17
 <211> 1854
 <212> DNA
 <213> Homo sapiens

<400> 17

gtctagttag ggacagacca agcagcaaaa acaaattgca atataatgtg ataagttctt 60
 taaaagaggt aagagcaacy tgctttggga gcagagaaga gggagaaaag agcatcttgc 120
 ctggatgagc caggggacac agaagagaag cccactatct catttaatct ttacaactct 180

```

cttgcaaggt tccctgggtt gtgaaaatac atgagataaa tcatgaaggc cactatcatc 240
ctccttctgc ttgcacaagt ttcctggggc tggaccgttt caacagagag gcttatttga 300
ctttatgcta ggaagatgag gcttctgggg ataggccag aagttcctga tgaccgcgac 360
ttcgagcccc tccctagggc ccagtgtgcc ccttcgctg tcaatgccat cttcgagtgg 420
tccagtgttc tgatttgggt ctggacaaag tgccaaagga tcttccccct gacacaactc 480
tgctagacct gcaaaacaac aaaataaccg aaatcaaaga tggagacttt aagaacctga 540
agaaccttca cgcatgtatt cttgtcaaca ataaaattag gcaaagttag tctggggagc 600
atttacacct ttggtgaaag ttggaacgac tttatctgtc caagaatcag ctgaaggaat 660
tgccagaaaa aatgccc aaa actcttcagg agctgcgtgc ccatgagaat gagatcacca 720
aagtgcgaaa agttactttc aatggactga accagatgat tgtcatagga actgggcacc 780
aatccgctga agagctcagg aattgaaaat ggggctttcc agggaatgaa ggaagctctc 840
ctacatccgc attgctgata ccaatatcac cagcattcct caaggtcttc ctccttcct 900
tacgggaatt acatcttgat ggcaacaaaa tcagcagagt tgatgcagct agcctgaaag 960
gactgaataa ttgggctaag ttgggattga gtttcaacag catctctgct gttgacaatg 1020
gctctctggc caacacgct catctgaggg agcttcactt ggacaacaac aagcttacca 1080
gagtacctgg tgggctggca gagcataagt acatccaggt tgtctactt cataacaaca 1140
atatctctgt agttggatca agtgacttct gcccacctgg acacaacacc aaaaaggctt 1200
cttattcggg tgtgagtctt ttcagcaacc cggtcagta ctgggagata cagccatcca 1260
ccttcagatg tgtctacgtg cgctctgcca ttcaactcgg aaactataag taattctcaa 1320
gaaagccctc atttttataa cctggcaaaa tcttgtaaat gtcattgcta aaaaataaat 1380
aaaagctaga tactggaaac ctaactgcaa tgtggatgtt ttaccacat gacttattat 1440
gcataaagcc aaatttccag tttaagtaat tgccataaat aaaaagaaat ttgctctgcc 1500
attttcagaa tcatcttttg aagctttctg ttgatgttaa ctgagctact agagatattc 1560
ttatttcact aaatgtaaaa tttggagtaa atatatatgt caatatttag taaagctttt 1620
cttttttaat ttccaggaaa aaataaaaa agtatgagtc ttctgtaatt cattgagcag 1680
ttagctcatt tgagataaag tcaaatgcca aacactagct ctgtattaat ccccatcatt 1740
actggtaaag cctcatttga atgtgtgaat tcaatacagg ctatgtaaaa tttttactaa 1800
tgtcattatt ttgaaaaaat aaatttaaaa atacattcaa aattaaaaaa aaaa 1854

```

```

<210>      18
<211>     1585
<212>      DNA
<213>     Homo sapiens

<400>      18

```

```

gattcggcac gatggaatcc accagctaca tccagctccc tgaggcagag ttgagaatgg 60
agagaatgtt acctctcctg actctggggc tcttgccggc tgggttctgc cctgctgtcc 120
tctgccaccc taacagccca cttgacgagg agaatctgac ccagggagaa ccaagaccga 180
gggacacacg tggacctcgg attagcctcc gccaacgtgg gacttcgctt tcagcctgta 240
caagcagtta gtctgaaag gccctgata agaatgtcat cttctcccca ctgaggcatc 300
tccaccgctt tggccttcct gtctctgggg ggcccataat accaccctgg acagagattc 360
tcaaaggcct caagttcaac ctcacggaga cttctgaggc agaaattcac cagagctttc 420
cagcacctcc tgcgcaccct caatcagtc agcgatgagc tgcaagctga gtatgggaaa 480
tgccatgttt gtcaaagagc aactcagtc gctggacagg ttcacggagg atgccaagag 540
gctgtatggc tccgaggcct ttgccactga ctttcaggac tcagctgcag ctaagaagct 600
catcaacgac tacgtgaaga atggaactag ggggaaaatc acagatctga tcaaggacct 660
tgactcgcag acaatgatgg tcctggtgaa ttacatcttc tttaaagcca aatgggagat 720
gccctttgac cccaagata ctcacagtc aaggttctac ttgagcaaga aaaagtgggt 780
aatggtgccc atgatgagtt tgcacacct gactatacct tacttccggg acgaggagct 840
gtcctgcacc gtggtggagc tgaagtacac aggcaatgcc agcgcaactc tcatcctccc 900
tgatcaagac aagatggagg aagtggaagc catgctgctc ccagagacc cgaagcggtg 960
gagagactct ctgagttca gagagatagg tgagctctac ctgccaaagt tttccatctc 1020
gagggactat aacctgaacg acatacttct ccagctgggc attgaggaag ccttcaccag 1080
caaggctgac ctgtcagggc tcacaggggc caggaaaccta gcagtctccc aggtggtcca 1140
taaggctgtg cttgatgtat ttgaggaggg cacagaagca tctgctgcca cagcagtcaa 1200
aatcaccttc ctttctgcat tagtgagac aaggaccatt gtgcgtttca acagggcctt 1260
cctgatgatc attgtccctt acagacacc agaacatctt cttcatgagc aaagtcacca 1320
atcccaagca agcctagagc ttgccatcaa gcagtggggc tctcagtaag gaacttgaa 1380
tgcaagctgg atgcctgggt ctctgggcac agcctggccc ctgtgcaccg agtggccatg 1440
gcattgtgtg ccctgtctgc ttatccttgg aaggtgacag cgattccctg tgtagctctc 1500
acatgcacag gggcccatgg actcttcagt ctggagggtc ctgggcctcc tgacagcaat 1560
aaataatttc gttggacacg ttaaa
1585

```

```

<210>      19
<211>      1390
<212>      DNA
<213>      Homo sapiens

<400>      19

```

```

ggcaccacca ctaacctggg acagtgaatc gacaatgccg tcttctgtct cgtggggcat 60

```

cctcctgctg gcaggcctgt gctgcctggt ccctgtctcc ctggctgagg atccccaggg 120
 agatgctgcc cagaagacag atacatccca ccatgatcag gatcacccaa ccttcaacaa 180
 gatcaccccc aacctggctg agttcgctt cagcctatac cgccagctgg cacaccagtc 240
 caacagcacc aatatcttct tctccccagt gagcatcgct acagcctttg caatgctctc 300
 cctgggggac caaggctgac actcacgatg aaatcctgga gggcctgaat ttcaacctca 360
 cggagattcc ggaggctcag atccatgaag gcttccagga actcctccgt accctcaacc 420
 agccagacag ccagctccag ctgaccaccg gcaatggcct gttcctcagc gagggcctga 480
 agctagtgga taagtctttg gaggatgtta aaaagtgtga ccactcagaa gccttcactg 540
 tcaacttcgg ggacaccgaa gaggccaaga aacagatcaa cgattacgtg gagaagggtg 600
 ctcaagggaa aattgtggat ttggtcaagg agcttgacag agacacagtt ttgctctgg 660
 tgaattacat cttctttaa ggcaaatggg agagaccctt tgaagtcaag gacaccgagg 720
 aagaggactt ccacgtggac caggtgacca cgtgaagggt gcctatgatg aagcgttag 780
 gcatgtttaa catccagcac tgtaagaagc tgtccagctg ggtgctgctg atgaaatacc 840
 tggggcaatg ccaccgcat cttcttctg cctgatgagg ggaaactaca gcacctggaa 900
 aatgaactca cccagatat catcaccaag ttcttgaaaa atgaagacag aaggtctgcc 960
 agcttacatt taccctaaact gtccattact ggaacctatg atctgaagag cgtcctgggt 1020
 caactgggca tactaagggt cttcagcaat ggggctgacc tctccgggggt cacagaggag 1080
 gcacctctga agctctccaa ggccgtgcat aaggctgtgc tgaccatcga cgagaaaggg 1140
 actgaagctg ctggggccat gtttttagag gccataccca tgtctatccc ccccgaggtc 1200
 aagttcaaca aaccttttgt cttcttaatg attgaacaaa ataccaagtc tcccctcttc 1260
 atgggaaaag tggatgaatc caccacaaaa taactgcctc tcgctcctca acccctcccc 1320
 tccatccctg gccccctccc tggatgacat taaagaaggg ttgagctgggt ccctgcctgc 1380
 atgtgactgt 1390

<210> 20
 <211> 1534
 <212> DNA
 <213> Homo sapiens
 <400> 20

ggaagatccc aacagtttgc gccataaata taactttatc gcggacgtgg tggagaagat 60
 cggccctgcc gtggttcata tcgaattgtt tcgcaagett ccgttttcta aacgagagggt 120
 gccgggtggct agtgggtctg ggtttattgt gtcggaagat ggactgatcg tgacaaatgc 180
 ccacgtgggtg accaacaagc accgggtcaa agttgagctg aagaacgggtg ccacttacga 240
 agccaaaatc aaggatgtgg atgagaaagc agacatcgca ctcacaaaa ttgaccacca 300

18

gggcaagctg cctgtectgc tgcttggcgg ctccctcagag ctgcgggcgg gagagtctgt 360
 ggtcgccatc ggaagcccg tttcccttca aaacacagtc accaccggga tctgtgagcac 420
 caccagcga ggcggcaaag agctggggct ccgcaactca gacatggact acatccagac 480
 cgacgccatc atcaactatg ggaaactccg ggaggcccg tagtaaacct ggacggtgaa 540
 gtgattggaa ttaacacttt gaaagtgaac gctggaatct cctttgcaat cccatctgat 600
 aagattaaaa agttcctcac ggagtcccat gaccgacagg ccaaaggaaa agccatcacc 660
 aagaagaagt atattggtat ccgaatgatg tcactcacgt ccagcaaagc caaagagctg 720
 aaggaccggc accgggactt cccagacgtg atctcaggag cgtatataat tgaagtaatt 780
 cctgataccc cagcagaagc tgggtgggtct caaggaaaac gacgtcataa tcagcatcaa 840
 tggacagtcc gtggtctccg ccaatgatgt cagcgacgtt cattaaaagg gaaagcacc 900
 tgaatggtt ggtccgagg ggtaatgaag atactcatgat cacagtgaatt cccgaagaaa 960
 ttgaccata ggcagaggca tgagctggac ttcattgttc cctcaaagac tctcccgtag 1020
 gatgacggat gaggactctg ggctgctgga ataggacact caagactttt gactgccatt 1080
 ttgtttgttc agtggagact ccctggccaa cagaatcctt cttgatagtt tgcaggcaaa 1140
 acaaatgtaa tgttgagat ccgcaggcag aagctctgcc ccttctgtat cctatgtatg 1200
 cagtgtgctt tttcttgcca gcttgggcca tttctgctta gacagtgcgc atttgtctcc 1260
 tcctttaact gaggcatcat cttagtccaa ctaatgcagt cgatacaatg ccgtagatag 1320
 aagaagcccc acgggagcca ggatgggact ggtcgtgttc gtgcttttct ccaagtcagc 1380
 acccaaaggc caatgcacag agaccccggt tgggtgagcg ctggcttctc aaacggccga 1440
 agttgcctct ttttagaatc tctttggaat tgggagcacg atgactctga gtttgagcta 1500
 ttaaagtact tcttacacat tgaaaaaaaa aaaa 1534

<210> 21
 <211> 2559
 <212> DNA
 <213> Homo sapiens
 <223> unsure at all n locations
 <400> 21

agctgtcggg gcggttagtt cgatttcgag ctgagggttt ccccgccgc cagggtgnact 60
 tctcatcgct tgtttttctt tttgatttt tcttccacc gccgttgccg ccttccccgt 120
 cctggccgtc cgccctccgc cctctgcagg gacatctcta caccgttccc atccgggaac 180
 agggcaacat ctacaagccc aacaacaagg cctggcaga cgagctgagc gagaagcaag 240
 tgtacgacgc gcacaccaag gagatcgacc tggccaaccg cgaccctaaa cactcaacg 300
 atgacgtggt caagattgac tttgaagatg tgattgcaga accagaaggg acacacagtt 360

ttgacggcat ttggaaggc cagcttcacc accttcactg tgacgaaata ctggttttac 420
 cgcttgctgt ctgccctctt tggcatcccg atggcactca tctggggcat ttacttcgcc 480
 attctctctt tctgcacat ctgggcagtt gtaccatgca ttaagagctt cctgattgag 540
 attcagtgca tcagccgtgt ctattccatc tacgtccaca ccgtctgtga cccactcttt 600
 gaagctgttg ggaaaatatt cagcaatgtc cgcacaaact tgcagaaaga aatataaatg 660
 acatttcaag gatagaagta tacctgattt tttttcttt taattttctt ggtgccaatt 720
 tcaagttcca agttgcta acagcaacaa tttatgaatt gaattatctt ggttgaaaat 780
 aaaaagatca ctttctcagt tttcataagt attatgtctc ttctgagcta tttcatctat 840
 ttttggcagt ctgaattttt aaaaccatt taaattttt tccttacctt tttatttgca 900
 tgtggatcaa ccatcgcttt attggtgag atatgaacat attgttgaaa ggtaatttga 960
 gagaaatag aagaactgag gaggaaaaaa aaaaaaaga aaagaaccaa caacctcaac 1020
 tgcctactcc aaaaatgttg tcatcttatg ttaaggaag aattccaggg tatggccatg 1080
 gagtgtaaa gtatgtgggc agattttcag caaactcttt tccactgtt taaggagtta 1140
 gtggattact gccattcact tcataatcca gtaggatcca gtgatcctta caagttagaa 1200
 aacataatct tctgcctct catgatccaa ctaatgcctt actcttcttg aaattttaac 1260
 ctatgatatt ttctgtgctt gaattttgt tatgtagaa acaagacctc agtgccttcc 1320
 tgtttttcc attttccttt tcaaataggg tctaactcag caactcgctt taggtcagca 1380
 gcctccctga agaccaaagt tagaatatcc atgacctagt ttccatgag tgtttctgac 1440
 tctgagctac agagtctggt gaagctcact tctgggcttc atctggcaac atctttatcc 1500
 gtagtgggta tgggtgacac tagcccaatg aaatgaatta aagtgggacc aatagggctg 1560
 agctctctgt gggctgggca gtccctggga gccagcttcc cctgcctctc atcaactgaa 1620
 tgaggtcagc atgtctatc agcttcgttt attttcaaga ataactcacg tttcctgaat 1680
 ccaaactaat ccatcaccgg ggtggtttag tggctcaaca ttgtgttccc atttcagctg 1740
 atcagtgggc ctccaaggag gggctgtaa atggaggcca ttgtgtgagc ctatcagagt 1800
 tgctgcaaac ctgaccctg ctcatgaaag cacttgcaac cgtctgttat gctgtgacac 1860
 atggccctcc cccctgccag gagctttgga cctaatacaa gcatctcttt gccagaaag 1920
 aagatggggg agggaggcgt aataaaaaga ttgaagtatt ttgctggaat aagttcaaat 1980
 tcttctgaac tcaaactgag gaatttcacc tgtaaacctg agtcgtacag aaagctgcct 2040
 ggtatatcca aaagcttttt attctctctg ctcatattgt gattctgcct ttggggactt 2100
 ttcttaaacc ttcagttatg attttttttt catacactta ttggaactct gcttgatttt 2160
 tgcctcttcc agtcttctg acactttaat taccaacctg ttacctactt tgactttttg 2220
 catttaaac agacactggc atggatatag ttttactttt aaactgtgta cataactgaa 2280

20

aatgtgctat actgcatact ttttaaagt aaagatattt ttatctttat atgaagaaaa 2340
 tcacttagga aatggctttg tgattcaatc tgtaaactgt gtattccaag acatgtctgt 2400
 tctacataga tgcttagtcc ctcatgcaaa tcaattactg gtccaaaaga ttgctgaaat 2460
 tttatatgct tactgatata ttttacaatt ttttatcatg catgtcctgt aaagggtaca 2520
 agcctgcaca ataaaaatgt ttaacgggta aaaaaaaaaa 2559

<210> 22
 <211> 981
 <212> DNA
 <213> Homo sapiens

<400> 22
 gcggagtctc caactgggag agctgcagct gccgagagga ggagaacgct gaggtcggtc 60
 ggaccaacgg acgcgctgac cgctgccaac tgcagctcgc gctgcctcct gctcgcgccg 120
 tgccactaag gtagtccgcc tttctatgag cctccccaa gattagctgg gtgcgggggtg 180
 gtgggagccg ttctttggtg gctgaagccc ctctcctgct gctcctcctg caggtcactc 240
 ccgcctccga gagccagag ccgagatgga aacgggccag gagctgatcc ccctggccaa 300
 ggagatgatg gcccagaagc gcaaggggaa gatggtgaag ctgtacgtgc tggggcagcg 360
 tgctggcctt cttcggcgtg gtgctcggcc tgatggagac tgtgtgcagc cccttcacgg 420
 ccgccagacg tctgcgggac caggaggcag ccgtggcgga gctgcaggcc gccctggagc 480
 gacaggctct ccagaagcaa gccctgcagg agaaaggcaa gcagcaggac acggtcctcg 540
 gcggccgggc cctgtccaac cggcagcacg cctcctagga actgtgggag accagcggag 600
 tgggagggag acgcagtaga cagagacaga ccgagaagga agggagagac agagggggcg 660
 cgcgcacagg agcctgactc cgctgggaga gtgcaggagc acgtgctgtt ttttatttgg 720
 acttaacttc agagaaaccg ctgacatcta gaactgacct accacaagca tccaccaaag 780
 gagtttggga ttgagttttg ctgctgtgca gcactgcatt gtcatgacat ttccaacact 840
 gtgtgaatta tctaaatgcg tctaccattt tgcactaggg aggaaggata aatgcttttt 900
 atgttattat tattaattat tacaatgacc accattttgc attttgaaat aaaaaacttt 960
 ttataccaaa aaaaaaaaaa a 981

<210> 23
 <211> 835
 <212> DNA
 <213> Homo sapiens

<400> 23
 gcactcccaa agaactgggt actcaacact gaggcagatc tgttctttga ggctaaaaac 60
 catgtgctgt accaagagtt tgctcctggg ctgctttgat gtcagtgtg ctactccacc 120

21

tctgcggcga atcagaagca gcaagcaact ttgactgctg tcttgggata cacagaccgt 180
 attcttcata ctaaatttat tgtgggcttc acacggcagc tggccaatga aggctgtgac 240
 atcaatgcta tcatacttca cacaagaaa aagttgtctg tgtgcgcaaa tccaaaacag 300
 acttgggtga aatatattgt gcgtctcttc agtaaaaaag tcaagaacat gtaaaaactg 360
 tggcttttct ggaatggaat tggacatagc ccaagaacag aaagaacctt gctgggggtg 420
 gaggtttcac ttgcacatca tggaggggtt agtgcttata taatttgtgc ctactggac 480
 ttgtccaatt aatgaagtg attcatattg catcatagtt tgctttgttt aagcatcaca 540
 ttaaagttaa actgtatttt atgttattta tagctgtagg ttttctgtgt ttagctattt 600
 aatactaatt ttccataagc tattttggtt tagtgcaaag tataaaatta tatttggggg 660
 ggaataagat tatatggact ttcttgcaag caacaagcta ttttttaaaa aaaactattt 720
 aacattcttt tgtttatatt gttttgtctc ctaaattggt gtaattgcat tataaaataa 780
 gaaaaatatt aataagacaa atattgaaaa taaagaaaca aaaagttcaa aaaaa 835

<210> 24
 <211> 981
 <212> DNA
 <213> Homo sapiens

<400> 24

ggcggccgga gagctcttgc gcgtcttgtt cttgcctggt gtcgggtggt agtttctgag 60
 acttgtgttg ggactgctga taggaagatg tcttcaggaa atgctaaaat tgggcaccct 120
 gcccccaact tcaaaagccac agctgttatg ccagatgggc agtttaaaga tatcagcctg 180
 tctgactaca aaggaaaata tgttgtgttc ttcttttacc ctcttgactt cacctttgtg 240
 tgccccacgg agatcattgc ttttcagtga tagggcagaa gaatttaaga aactcaactg 300
 ccaagtgatt ggtgcttctg tgggattctc acttctgtca tctagcatgg ggtcaataca 360
 cctaagaac aaggaggact gggacccatg aacattcctt tggatcaga cccgaagcgc 420
 accattgctc aggattatgg ggtcttaag gctgatgaag gcatctcgtt caggggcctt 480
 tttatcattg atgataaggg tattcttcgg cagatcactg taaatgacct cctgtttggc 540
 cgctctgtgg atgagacttt gagactagt caggccttcc agttcactga caaacatggg 600
 gaagtgtgcc cagctggctg gaaacctggc agtgatacca tcaagcctga tgtccaaaag 660
 agcaaagaat atttctccaa gcagaagtga gcgctgggct gttttagtgc caggctgcgg 720
 tgggcagcca tgagaacaaa acctcttctg tatttttttt ttccattagt aaacacaaag 780
 acttcagatt cagccgaatt gtggtgtctt acaaggcagg cctttcctac aggggggtgga 840
 gagaccagcc tttcttcctt tggtaggaat ggcctgagtt ggcgttgtgg gcaggctact 900
 gggttgtatg atgtattagt agagcaaccc attaatcttt ttagtttgt attaaacttg 960

aactgagaaa aaaaaaaaaa a

981

<210> 25
<211> 1642
<212> DNA
<213> Homo sapiens

<400> 25

gaaaaaggcg agcccgcccc ccctggagac cccggtctca cggagttgac gtcacgacct 60
acgtgagggg gacctgcggg tgctgcgact gtgagaagcg ctgtggcgcc ctggacgtgg 120
tcttcgtcat cgacagctcc gagagcattg ggtacaccaa cttcacactg gagaagaact 180
tcgtcatcaa cgtggtcaac aggcctgggtg ccatcgctaa ggacccaag tccgagacag 240
ggacgcgtgt gggcgtggtg cagtacagcc acgagggcac ctttgaggcc atccagctgg 300
acgacgaaca tatcgactcc ctgtcgagct tcaaggaggc tgtcaagaac ctcgagtggg 360
ttgcggggcg cacctggaca ccctcagccc tcaagtttgc ctacgaccgc ctcatcaagg 420
agagccggcg ccagaagaca cgtgtgtttg cgggtggtcat cacggacggg cgccacgacc 480
ctcgggacga tgacctcaac ttgcggggcg tgtgcgaccg cgacgtcaca gtgacggcca 540
tcggcatcgg ggacatgttc cagcagaagc acgagagtga aaacctctac tccatcgctt 600
gcgacaagcc acagcaggtg cgcaacatga cgtgtttctc ccgacctggt cggttgagaa 660
gttcatcgat gacatgggag gacgtcctct gcccggaccc tcagatcgtg tgcccagacc 720
ttccctgcca aacagagctg tccgtggcac agtcacgca gcggcccggt gacatcgtct 780
tcctgctgga cggctccgag cggctgggtg agcagaactt ccacaaggcc cggcgcttcg 840
tggagcaggt ggcgcgcgcg ctgacgctgg cccggaggga cgacgaccct ctcaacgcac 900
gcgtggcgct gctgcagttt ggtggccccg gcgagcagca ggtggccttc ccgctgagcc 960
acaacctcac ggccatccac gaggcgctgg agaccacaca atacctgaac tccttctcgc 1020
acgtggggcg aggcgtgggt caccgcatca atgcatcgt gcgcagccag cgtggcgggc 1080
ggcgagggca cgcagagctg tccttcgtgt tcctcacgga cggcgtcacg ggcaacgaca 1140
gtctgcacga gtcggcgcac tccatgcgca agcagaacgt ggtacccacc gtgctggcct 1200
tgggcagcga cgtggacatg gacgtgctca ccacgctcag cctgggtgac cgtgccgccc 1260
tgttccacga gaaggactat gacagcctgg cgcaaccggg cttcttcgac cgcttcatcc 1320
gctggatctg ctacgcccgc cgccggggcc ccgcagtcga gggctcgtgag cccaccccg 1380
ccatggtgct aagcggggcc ggggtcccaca cggccagcac cgctgctcac tggagcagc 1440
ccctgggcct gcacctctcc agctcctccc acgggggtccc cgtagccccg gcccccgccc 1500
agccccaggt ctccccaggc cctccgcagg ctgcccggcc tccctccccc tgcagccatc 1560
ccaaggctcc tgacctacct ggccccctgag ctctggagca agccctgacc caataaaggc 1620

23

tttgaaccca aaaaaaaaaa aa

1642

<210> 26
 <211> 163
 <212> DNA
 <213> Homo sapiens

<400> 26

gaccagtttg tcaagaaggg tagctgctgg agggggacac accctctgtc tgatccctta 60

tcaaagagga caaggaaact atagagctga ttttagaata ttttacaat acatgccttc 120

cattggaatg ctaagatttt ctactgcttc tggggacggg aaa 163

<210> 27
 <211> 1746
 <212> DNA
 <213> Homo sapiens

<223> unsure at all n locations

<400> 27

cagcgctccc actctcgcc gacacccctc atggccaacc gttacaccat ggatctgact 60

gccatctacg agagcctcct gtcgctgagc cctgacgtgc ccgtgccatc cgaccatgga 120

gggactgagt ccagcccagg ctggggctcc tcgggaccct ggagcctgag cccctccgac 180

tccagcccgt ctggggtcac ctcccgcctg cctggccgct ccaccagcct agtggaggggc 240

cgcagctgtg gctgggtgcc cccacccctt ggcttcgcac cgctggctcc ccgcctgggc 300

cctgagctgt caccctcacc cacttcgccc actgcaacct ccaccacccc ctgcgctac 360

aagactgagc tatgtcggac cttctcagag agtgggctg gctgctacgg ggccaagtgc 420

cagtttgccc atggcctggg cgagctgcgc caggccaatc gccaccccaa atacaagacg 480

gaactctgtc acaagttcta cctccagggc cgctgccctt acggctctcg ctgccacttc 540

atccacaacc ctagegaaga cctggcgggc ccggggccacc ctctctgtgt tcgccagagc 600

atcagcttct ccggcctgcc ctctggccgc cggacctcac caccaccacc aggcctggcc 660

ggcccttccc tgtcctccag ctccctctcg cctccagct cccaccacc acctggggac 720

cttccactgt nacctctgc cttctctgt gcccctggca cccctctggc tcgaagagac 780

cccacccag tctgttgccc ctctgccga agggccactc ctatcagcgt ctggggggccc 840

ttgggtggcc tgggttcggac cccctctgta cagtccctgg ggatccgacc ctgatgaata 900

tgccagcagc ggcagcagcc tggggggctc tgactctccc gtcttcgagg cgggagtttt 960

tgcaccacct cagcccgtgg cagccccccg gcgactcccc atcttcaatc gcattctctgt 1020

ttctgagtga caaagtgact gcccggtcag atcagctgga tctcagcggg gagccacgtc 1080

tcttgactg tggctctctgc atggacccca gggctgtggg gacttggggg acagtaatca 1140

agtaatcccc ttttccagaa tgcattaacc cactcccctg acctcacgct ggggcaggtc 1200

cccaagtgtg caagctcagt attcatgatg gtgggggatg gagtgtcttc cgaggttctt 1260
 gggggaaaaa aaattgtagc atatttaagg gaggaatga accctctccc ccacctcttc 1320
 cctgccccaa tctgtctcct agaattctat gtgctgtgaa taataggcct tcaactgcccc 1380
 tccagttttt atagacctga ggttccagtg tctcctggta actggaacct ctctgaggg 1440
 ggaatctctg tgctcaaatt accctccaaa agcaagtagc caaagccgtt gccaaacccc 1500
 acccataaat caatggggcc tttatttatg acgactctat ttattctaata atgattttat 1560
 agtatattata tatattgggt cgtctgcttc ccttgtaatt ttcttctttt ttttgtaata 1620
 ttgaaaacga cgatataatt attataagta gactataata tatttagtaa tatatattat 1680
 taccttaaaa gtctattttt gtgttttggg cttttttaa taaacaatct gagtgtaaaa 1740
 aaaaaa 1746

<210> 28
 <211> 1884
 <212> DNA
 <213> Homo sapiens

<400> 28
 cgtcgtagcc ccaacctcga cggctgcccgt gggcccggtc gcgtctgcct tggagaagaa 60
 gacaaagagc aagggggccct acatctgcgc tctgtgcgcc aaggagtcca agaacggcta 120
 caatctccgg aggcacgaag ccatccacac gggagccaag gccggccggg tcccctcggg 180
 tgctatgaag atgccgacca tgggtgcccc gagcctctcg agcgtgcccc agctgagcgg 240
 agccggcggg ggagggggag aggcgggtgc cggcggcggc gctgcccag tggccggcgg 300
 tggcgtgggt accacgaccg cctcggggaa gcgcacccg aagaaccatg cctgcgagat 360
 gtgtggcaag gccttccgag acgtctacca cctgaaccga cacaagctgt cgcactcggg 420
 cgagaagccc taccagtgcc cgggtgtgca gcagcgcttc aagcgcaagg accgcatgag 480
 ctaccacgtg cgtctacatg acggcgctgt ggcacaagccc tacaactgct cccactgtgg 540
 caagagcttc tcccggccgg atcacctcaa tcttcacgtc agacaagtgc actcaacaga 600
 acggcccttc aaatgtgaga aatgtgaggc aactttcgcc acgaaggatc ggctgcgggc 660
 gcacacagta cgacacgagg agaaagtgc atgtcacgtg tgtggcaaga tgctgagctc 720
 ggcttatatt tcggaccaca tgaagggtga cagccagggt cctcaccatg tctgtgagct 780
 ctgcaacaaa ggtactgggt aggtttgtcc aatggcggcg gcagcggcag cggccgggca 840
 gcggcagcag cggcagcagt agcagccccct cccacagctg tgggctccct ctcgggggcg 900
 gaggggggtc ctgtgagctc tcagccactt ccttcccaac cctggtgagc tccaagtgg 960
 ttgcggggga gaggggagaa tggagtagag tcccttggta caagctctc tccccctct 1020
 tttccacca actctattt cctaccaac caaggagcct ccagaaggaa aggaggaaga 1080

25

aatgttttct taggggaatt cgctagggtt taacgatttg tttctcctgc tectcttcta 1140
 tcagacctga cccacacaaa acctgtcccc tcggtttgtg tgaagtcccc tggacagtgg 1200
 gcaggggtgg cagaggacac gagcagccac tgcccgtagc cctctcctc tctgtaagcc 1260
 catgcctctg ctcccaggg acttgtagc ctctccctc gacggtcctc ttctctcctt 1320
 ccagtctct cccctgctg tctgcagccc ctccccggg agttgggtgct ttcttttctt 1380
 tttttttttt tttccaggg gagggaggag aggaaggagg gggatcagag ctgtcccaaa 1440
 gagggaaaag ggtgaggttt gaggggggc agaagcaggc ccggcaaagg ttgtaccttc 1500
 ataaggtggt atgggggggt ggggtcaggc cctgaacatc gtcctacttg agaactctgc 1560
 aggggaaaaa gtcaagggga gcaggaggaa gagccaggag gccagaggca gagaagagat 1620
 ggagtcttag gggccaggg gagcgagggg tccagggcct agagggtgctt cctgggggag 1680
 ggggaatgca gccagtgtcc cctccccctc tccacccca gtcacagccc tgggtcttgc 1740
 ttttcatccc tcttccccac gacagaagaa gttgtggccc tggccatgtc atcgtgttcc 1800
 tgtgtccctt gcatgtacct caccctccac ccttctctt tgcgcggacc ccattacaat 1860
 aaattttaaa taaaatcctg aaaa 1884

<210> 29
 <211> 1563
 <212> DNA
 <213> Homo sapiens
 <400> 29

tcacctccag gatacagaca gcccccttc agcccagccc agccaggtct cctacaccgc 60
 caccatgcc a ttcggttaaca cccacaacaa gttcaagctg aattacaagc ctgaggagga 120
 gtaccccgac ctacgcaaac ataacaacca catggccaag gtactgacct ttgaactcta 180
 caagaagctg cgggacaagg agactccatc tggcttcaat gtagacgatg tcatccagac 240
 aggagtggac aaccaggtc accccttcat catgaccgtg ggctgcgtgg ctggtgatga 300
 ggagtcttac gaagttttca aggaactctt tgaccccatc atctcgatc gccacggggg 360
 ctacaaaccc acttgacaag cacaagactg acctcaacca ttgaaaacct caagggtgga 420
 gacgacctgg accctaacta cgtgctcagc agccgcgtcc gactggccg cagcatcaag 480
 ggctacacgt tgccccaca ctgctccctg ggcgagcgcc gggcggtgga gaagctctct 540
 gtggaagctc tcaacagcct gacgggagag ttcaaaggga agtactacct tctgaagagc 600
 atgacggaga aggagcagca gcagctcatc gatgaccact tcctgttcga caagcccggt 660
 tccccgctgc tgctggcctc aggcattggc cgcgactggc ccgacgccc tggatctggc 720
 acaatgacaa caagagcttc ctggtgtggg tgaacgagga ggatcacctc cgggtcatct 780
 ccattggagaa ggggggcaac atgaaggagg ttttccgccc cttctgcgta gggctgcaga 840

26

```

agattgagga gatctttaag aaagctggcc accccttcat gtggaaccag cacctgggct 900
acgtgctcac ctgcccatcc aacctgggca cctgggctgc gtggaggcgt gcatgtgaag 960
cctggcgcac ctgagcaagc accccaagtt cgaggagatc ctcacccgcc tgcgtctgca 1020
gaagaggggt acaggtggcg tggacacagc ctgccgtggg ctgagtattt gacgtgtcca 1080
acgtgatcgt gctgggctcg tccgaagtag aacaggtgca gctggtggtg gatggtgtga 1140
agctcatggt ggaaatggag aagaagttgg agaaaggcca gtccattgac gacatgatcc 1200
ccgcccagaa gtaggcgctt gccacactgc caccgactgc tggaaccag ccagtgggag 1260
ggcctggccc accagagtec tgctccctca ctctctgccc cgccccctgt cccagagtec 1320
cacctggggg ctctctccac ccttctcaga gttccagttt caaccagagt tccaaccaat 1380
gggtccatc ctctggattc tggccaatga aatatctccc tggcagggtc ctcttctttt 1440
cccagagctc caccccaacc aggagctcta gtaaatggag agctcccagc aactcggag 1500
cttgtgcttt gtctccacgc aaagcgataa ataaaagcat tgggtggcctt aaaaaaaaaa 1560
aaa
1563

```

```

<210>      30
<211>      2263
<212>      DNA
<213>      Homo sapiens
<223>      unsure at all n locations
<400>      30

```

```

ctcgagacaa gcccgatatgt gtcaacacct atggaagcta cagggtgccgg accaacaaga 60
agtgcagtcg gggctacgag cccaacgagg atggcacagc ctgcgtgggg actctcggcc 120
agtcaccggg cccccgccc accnnnnna cncggggac cggggctggg agcaagcagg 180
cggcgcgccc gggggcagag gcggcagcga gcgcccgtt cccacgccc taggcggcgg 240
ggccgagagc gggaggatgg ctccgagcgc tgaccccggc atgtccagga tgttacggt 300
cctgctgctg ctctggtttc tgcccatcac tgagggttcc cagcgggctg aacctatgtt 360
cactgcagtc accaactcag ttctgcctcc tgactatgac agtaatccca cccagctcaa 420
ctatggtgtg gcagttactg atgtggacca tgatggggac tttgagatcg tcgtggcggg 480
gtacaatgga cccaacctgg ttctgaagta tgaccgggcc cagaagcggc tggatgaacat 540
cgcggtcgat gagcgcagta acccctacta cgcgtgctcg gaccggcagg ggaacgcat 600
cggggtcaca gcctgcgaca tcgacgggga cggccgggag gagatctact tcctcaacac 660
caataatgcc ttctcggggg tggccacgta caccgacaag ttgttcaagt tccgcaataa 720
ccggtgggaa gacatcctga gcgatgaggt caacgtggcc cgtggtgtgg ccagcctctt 780
tgccggacgc tctgtggcct gtgtggacag aaagggtctt ggacgctact ctatctacat 840

```

27

tgccaattac gcctacggta atgtgggccc tgatgccctc attgaaatgg accctgaggc 900
 cagtgaacctc tcccggggca ttctggcgct cagagatgtg gctgctgagg ctggggtcag 960
 caaatataca gggggccgag gcgtcagcgt gggcccccac ctcagcagca gtgcctcggg 1020
 tatcttctgc gacaatgaga atgggcctaa cttccttttc cacaaccggg gcgatggcac 1080
 ctttgtggac gctgcggcca gtgctggtgt ggacgacccc caccagcatg ggcgaggtgt 1140
 cgccctggct gacttcaacc gtgatggcaa agtggacatc gtctatggca actggaatgg 1200
 cccccaccgc ctctatctgc aaatgagcac ccattgggaag gtccgcttcc gggacatcgc 1260
 ctacccaag ttctccatgc cctccctgt ccgcacggtc atcaccgccg actttgacaa 1320
 tgaccaggag ctggagatct tcttcaacaa cattgcctac cgcagctcct cagccaaccg 1380
 cctcttccgc gtcattcgta gagagcacgg agacccctc atcgaggagc tcaatcccgg 1440
 cgacgccttg gagcctgagg gccggggcac aggggggtgt gtgaccgact tcgacggaga 1500
 cgggatgctg gacctcatct tgtcccatgg agagtccatg gctcagccgc tgtccgtctt 1560
 ccggggcaat cagggttca acaacaactg gctgcgagtg gtgccaacgc acccggtttg 1620
 gggcctttgc caggggagct aaggctcgtc tctacaccaa gaagagtggg gccacactga 1680
 ggatcatcga cgggggctca ggctacctgt gtgagatgga gcccggtggc cactttggcc 1740
 tggggaagga tgaagccagc agtgtggagg tgacgtggcc agatggcaag atggtgagcc 1800
 ggaacgtggc cagcggggag atgaactcag tgctggagat cctctacccc cgggatgagg 1860
 acacacttca ggaccagcc ccaactggagt gtggccaagg attctcccag caggaaaatg 1920
 gccattgcca tggacaccaa tgaatgcac cagttcccat tcgtgtgccc tcgagacaag 1980
 ccggtatgtg tcaacaccta tggaagctac aggtgccgga ccaacaagaa gtgcagtcgg 2040
 ggctacgagc ccaacgagga tggcacagcc tgcgtggctc aagtggcctt tttagggtgg 2100
 tattcttcag ccgcctctag aatctctgag cctctctctc gggcctcata tctttctcta 2160
 ggcccttgac ttgccttca gttatatgca ctttaaatcc catcaataaa ggaaaaaaca 2220
 aaacaaaact aacagccttt gtggaaaact aaaaaaaaaa aaa 2263

<210> 31
 <211> 2310
 <212> DNA
 <213> Homo sapiens

<400> 31

cggcattcct cctgtagctg cacgaagcac cttggaagtt gttttcaacc atatccagcc 60
 tttgccgaat acatcctatc tgccacacat ccagcgtgag gtccctccag ctacaagggt 120
 ggcaccatgg cggagaagtt tgactgccac tactgcaggg atcccttgca ggggaagaag 180
 tatgtgcaaa aggatggcca ccaactgctg ctgaaatgct ttgacaagtt ctgtgccaac 240

acctgtgtgg aatgccgcaa gcccatcggg gcggactcca aggaggtgca ctataagaac 300
cgctttctggc atgacacctg cttccgctgt gcccaagtgc ttcacccctt gggccaatga 360
gaccttttggg gcccaaggaca acaagatcct gtgcaacaag tgcaccactc gggaggactc 420
ccccaaagtgc aaggggtgct tcaaggccat tgtggcagga gatcaaaacg tggagtacaa 480
ggggaccgtc tggcacaaaag actgcttcac ctgtagtaac tgcaagcaag tcatcgggac 540
tggaagcttc ttccttaaag gggaggactt ctactgctg acttgccatg agaccaagtt 600
tgccaagcat tgcgtgaagt gcaacaaggc catcacatct ggaggaatca cttaccagga 660
tcagccctgg catgccgatt gctttgtgtg tgttacctgc tctaagaagc tggctgggca 720
gcgtttcacc gctgtggagg accagtatta ctgctgggat tgctacaaga actttgtggc 780
caagaagtgt gctggatgca agaaccocat cactgggttt ggtaaaggct ccagtgtggt 840
ggcctatgaa ggacaatcct ggcacgacta ctgcttcac tgcaaaaaat gctccgtgaa 900
tctggccaac aagcgctttg tttccacca ggagcaagtg tattgtcccg actgtgcca 960
aaagctgtaa actgacaggg gtcctgtcc tgtaaaatgg catttgaatc tcgttctttg 1020
tgtccttact ttctgcccta taccatcaat aggggaagag tggctcctcc cttctttaa 1080
gttctccttc cgtcttttct cccattttac agtattactc aaataagggc acacagtgat 1140
catattagca tttagcaaaa agcaaccctg cagcaaagtg aatttctgtc cggctgcaat 1200
ttaaaaatga aaacttaggt agattgactc ttctgcatgt ttctcataga gcagaaaagt 1260
gctaatacatt tagccactta gtgatgtaag caagaagcat aggagataaa accccactg 1320
agatgcctct catgcctcag ctgggacca cccgtgtaga cacacgacat gcaagagttg 1380
cagcggctgc tccaactcac tgctcaccct cttctgtgag caggaaaaga accctactga 1440
catgcatggt ttaacttct catcagaact ctgcccttcc ttctgttctt ttgtgcttcc 1500
aaataactaa cacgaacttc cagaaaatta acatttgaac ttagctgtaa ttctaaactg 1560
acctttcccc gtactaacgt ttggtttccc cgtgtggcat gttttctgag cgttctact 1620
ttaagcatg gaacatcgag gtgatttggg aagtgtagaa agacctgaga aaacgagcct 1680
gtttcagagg aacatcgtca caacgaatac ttctggaagc ttaacaaaac taaccctgct 1740
gtccttttta ttgttttta ttaatatatt tgttttaatt gatagcaaaa tagtttatgg 1800
gtttggaaac ttgcatgaaa atattttagc cccctcagat gttcctgag tgctgaaatt 1860
catcctacag aagtaaccgc aaaactctag agggggagtt gagcaggcgc cagggctgtc 1920
atcaacatgg atatgacatt tcacaacagt gactagtga atcccttga acgtagtagt 1980
tgtctgctct ttgtccatgt gttaatgagg actgcaaagt ccttctgtt gtgattccta 2040
ggacttttcc tcaagaggaa atctggattt ccacctaccg cttacctgaa atgcaggatc 2100
acctacttac tgtattctac attattatat gacatagtat aatgagacaa tatcaaaagt 2160

29

aaacatgtaa tgacaataca tactaacatt cttgtaggag tggtagaga agctgatgcc 2220
tcattttctac attctgtcat tagctattat catctaacgt ttcagtgat ccttacagaa 2280
ataaagcagc atatgaataa aaaaaaaaaa 2310

<210> 32
<211> 3342
<212> DNA
<213> Homo sapiens

<400> 32

gaagaagtta agagcttcat ggatcgaag aagggattta cagaagtaa gtcgcagaat 60
ggagaattca tgaccacaa acttaaacat actgagaata ctttcagccg ccctggaggg 120
agggccagcg tggacaccaa ggaggctgag ggcgcccccc aggtggaagc cggcaaaagg 180
ctggaggagc ttcgtcgtcg tcgcggggag accgagagcg aagagttcga gaagctcaaa 240
cagaagcagc aggaggcggc tttggagctg gaggaactca agaaaaagag ggaggagaga 300
aggaaggtcc tggaggagga agagcagagg aggaagcagg aggaagccga tcgaaaactc 360
agagaggagg aagagaagag gagggctaaag gaagagattg aaaggcgaag agcagaagct 420
gctgagaaac gccagaagat gccagaagat ggcttgtcag atgacaagaa accattcaag 480
tgtttcactc ctaaaggttc atctctcaag atagaagagc gagcagaatt tttgaataag 540
tctgtgcaga aaagcagtgg tgtcaaatcg acccatcaag cagcaatagt ctccaagatt 600
gacagcagac tggagcagta taccagtga attgaggga caaaaagcgc aaacacctaca 660
aagccggcag cctcggtatc tcctgttctt gctgaagtg tacgcaacat caagagtatg 720
tgaggagaaag ggaatgtgtt ttcattcccc actgcagcag gcacaccaa taaggaaact 780
gcctggcttg aaggtagggg tttctagccg catcaatgaa tggctaacta aaacccaga 840
tggaacaag tcacctgctc ccaaaccctc tgacttgaga ccaggagacg tatccagcaa 900
gcggaacctc tgggaaaagc aatctgtgga taaggctact tccccccact aaggtttgag 960
acagttccag aaagaaccca agctcaagac gcaggacgag ctcagttgta gagggcta 1020
tcgctctgtt ttgtatttat gttgatttac taaattgggt tcattatctt ttatttttca 1080
atatcccagt aaacccatgt atattatcac taratttaat aatcacagtc tagagatgtt 1140
catggtaaaa gtactgcctt tgcacaggag cctgtttcta aagaaaccca tgctgtgaaa 1200
tagagacttt tctactgatc atcataactc tgtatctgag cagtgtatcc aaccacatct 1260
gaagtcaaca gaagatccaa gtttaaaatt gcctgcggaa tgtgtgcagt atctagaaaa 1320
atgaaccgta gtttttgttt ttttaaacac agaagtcag ttgtttctgc actttataat 1380
aaagcatgga agaaattatc ttagtaggca attgtaacac tttttgaaag taacccattt 1440
cagatttgaa atactgcaat aatggttgtc tttaaaaaaa aaaaagaaat gtactgttaa 1500

30

ggtattactt ttttcatgc tgatgattca tatctaaatt acattattat gttagctgac 1560
 agtggtagctg attttttagg ttggttggtt tgggatttc ttagtagtg atagtagcct 1620
 gaaccacatt ttagataact caattatgta tgtatgtgca tacacatata caaacacact 1680
 aatggtagaa tgctttttta tgtgctagac tattatattt agtagtatgt cattgtaact 1740
 agccaatatt acagcttttg aaaaattaaa aaatcacact atattaatat ttcataattg 1800
 ccaacagaaa catggcagat aggtatcaat atgttttcaa tgcctgatga cctataagaa 1860
 gaaagtattg aaaagaagag agattagaac tgctagaagg agttgaaatt ttctaaaaga 1920
 catagtattt agtttataat taaatgcatt cttgaagtc agtgtgaatt ttattaatgc 1980
 tatcatctcg accaagctca aagcctactt attagaaaca atgaagtca caatagggtca 2040
 taaggctctt tccttttcta aaattgaaag acaagaaatt tagtgccaat attgtacaga 2100
 cagaaattcc atgtatgagt ctcaacaaag actacctttg gctaaatgtc tagaagcaga 2160
 gaagtaagt gagcaaatc cagtgttgag gagtcagac agtacctga tctttatata 2220
 ctctgaagca tttcttcaa cttttctact tttatttgc attgatacct gtagtaagtt 2280
 gacaatgtgg tgaaatttca aaattatag taacttctac tagttttact ttctcccca 2340
 agtctttttt aactcatgat ttttacacac acaatccaga acttattata tagcctctaa 2400
 gtcttttttc ttcacagtag ataatgaaag agtctccag tgccttgga aaatgttcta 2460
 gtatagctgg atacatacag tggagtctta taaactcata cctcagtgga cttaaccaa 2520
 attgtgttag tctcaattcc taccacactg aggggagcct ccccaataa ctattttctt 2580
 atctgcagta ttcctccaga agagctaacc aggggcaggg ctggcatgag aagtgacatc 2640
 tgcgttacaa agtctatctt cctcataagt ctgtaagag caattgaatc ttctagcttt 2700
 agcaaaccta agccaaagga aggaaagcca cgaagaatgc agaagtcaaa ccctcatgac 2760
 aaagtaggca caagtctaca ataagctaaa tcagaattta caaatacaag tgtcccagg 2820
 agcattgact cccgtcattg gagtgaaatg gatcaaagt tgaattaagg cctatggtaa 2880
 ggtaacattg ctttggtgta cttttgaaca agagctcttc ctgatcacta ttacatattt 2940
 ttctagaaaa tctaaagttc agaagagaat gtatcacatgc tgacttttat tccaatattt 3000
 ggtaggagta agtttttagg tagaattttg ttcagtttgg atttaattct ttgaaaagta 3060
 aattccttgt ttactgggtt gactataatt ctctgttacc ttacagagg aaaaatgcaa 3120
 gctgactagc atgttctgtg aatctgccat tcttaaaaat ttataaaca cttgatactt 3180
 ttactgata atggatcgct ccaataaaca tatattgtga aaatgcattc acaataaatg 3240
 gaattccttc ctgcaaaatg tctttttctc acttattttt atgtacaata ttgatagtga 3300
 gaggtatgtc tattataata aagattatgg cacagcaaaa aa 3342

31

<211> 954
 <212> DNA
 <213> Homo sapiens

<400> 33

```

cagcctcaag attcacagca tctcagacgc agcctaggcc gcaccaggat gtcggacacc 60
gaggagcagg aatatgagga ggagcagccg gaagaggagg ctgcggttga ggaggaggaa 120
gccccgaag agcggagacc ggtggcagag ccagaagagg aacgccccaa accaagccgc 180
cccgtggtgc ctctttgat cccgcaaaag atcccagaag gggagcgcgt tgacttcgat 240
gacatccacc ggcaagcga tggagaaaga cctgctggag ctgcagacac tcatcgatgt 300
acatttcgag cagcggaaga aggaggaaga ggagctggtt gccttgaagg agcgattga 360
gcggcgccgg tcagagagag cccgagcaac agcgcttcag aactgagaag gaacgcgaac 420
gtcaggctaa gctggcggag gagaagatga ggaaggaaga ggaagaggcc aagaagcggg 480
cagaggatga tgccaagaaa aagaaggtgc tgtccaacat gggggcccat tttggcggct 540
acctggtcaa ggcagaacag aagcgtggtg agcggcagac ggggcgggag atgaaggtgc 600
gcatectctc cgagcgtaag aagcctctgg acattgacta catgggggag gaacagctcc 660
gggagaaagc ccaggagctg tcggactgga tccaccagct ggagtctgag aagttcgacc 720
tgatggcgaa gctgaaacag cagaaatatg agatcaacgt gctgtacaac cgcatcagcc 780
acgcccagaa gttccggaag ggggcaggga agggccgcgt tggaggccgc tgggaagtga 840
gatgccgccc cgacagtgg cacctgggaa gcctgggagt gtttgtccca tcggtagctt 900
gaaataaacg ctcccctcag acaccgctg ggttctctga tgttattatg gttg 954

```

<210> 34
 <211> 3183
 <212> DNA
 <213> Homo sapiens

<400> 34

```

gcgccgcacc tacaccagcc aaccagatc ccgaggtcg acagcgccc gcccagatcc 60
ccacgcctgc caggagcaag ccgagagcca gccggccggc gcactccgac tccgagcagt 120
ctctgtcctt cga'cccagc cccgcgccct ttccgggacc cctgccccgc gggcagcgt 180
gccaacctgc cggccatgga gaccccgtcc cagcggcgcg ccaccgcag cggggcgag 240
gccagctcca ctccgctgtc gccacccgc atcaccggc tgcaggagaa ggaggacctg 300
caggagctca atgatcgctt ggcggtctac atcgaccgtg tgcgctcgt ggaacggag 360
aacgcagggc tgcgccttcg catcaccgag tctgaagagg tggtcagccg cgagggtgcc 420
ggcatcaagg ccgcctacga ggccgagctc ggggatgccc gcaagaccct tgactcagta 480
gccaaggagc gcgccgcct gcagctggag ctgagcaaag tgcgtgagga gtttaaggag 540

```

ctgaaagcgc ggcaatacca agaaggaggg tgacctgata gctgctcagg ctcggtgaa 600
 ggacctggag gctctgctga actccaagga ggccgcactg agcactgctc tcagtggaga 660
 gcgcacgctg gaggggcagc tgcattatct gcggggccag gtggccaagc ttgaggcagc 720
 cctaggtgag gccaagaagc aacttcagga tgagatgctg cggggggtgg atgctggaga 780
 caggctgcag accatgaagg aggaactgga cttccagaag aacatctaca gtgaggagct 840
 gcgtgagacc aagcgccgtc atgagacccg actggtggag attgacaatg ggaagcagcg 900
 tgagtttgag agccggctgg cggatgcgct gcaggaactg cggggccagc atgaggacca 960
 ggtggagcag tataagaagg agctggagaa gacttattct gccaaactgg acaatgccag 1020
 gcagtctgct gagaggaaca gcaacctggg gggggctgcc cacgaggagc tgcagcagtc 1080
 gcgcattccg atcgacagcc tctctgccc gctcagccag ctccagaagc agctggcagc 1140
 caaggaggcg aagtttcgag acctggagga ctactggcc cgtgagcggg acaccagccg 1200
 gcggctgcct ggcggaagg gagcgggaga tggccgagat gcgggcaagg atgcagcagc 1260
 agctggacga gtaccaggag cttctggaca tcaagctggc cctggacatg gagatccacg 1320
 cctaccgcaa gctcttgag ggcgaggag agaggctacg cctgtcccc agccctacct 1380
 cgacgcgag ccgtggccgt gcttctctc actcatcca gacacaggtt gggggcagcg 1440
 tcacaaaaa gcgcaactg gagtccactg agagccgag cagcttctca cagcacgcac 1500
 gcactagcgg gcgctggggc cgtggaggag gtggatgagg agggcaagt tgcctggctg 1560
 cgacaacaag ccaatgagga ccagtcctg ggcaattggc agatcaagcg ccagaatgga 1620
 gatgatccct tgctgactta ccggttccca ccaaagtca cctgaaggc tgggcagggtg 1680
 gtgacgatct gggctgcagg agctggggc acccacagc cccctaccga cctggtgtgg 1740
 aaggcacaga acacctgggg ctgcgggaac agcctgcgta cggctctcat caactccact 1800
 ggggaagaag tggccatgcg caagctggtg cgtcagtgta ctgtggttga ggacgacgag 1860
 gatgaggatg gagatgacct gctccatcac caccacggct cccactgcag cagctcgggg 1920
 ggaccccgct gagtacaacc tgcgctcgcg caccgtgctg tgcgggacct gcgggcagcc 1980
 tgccgacaag gcatctgcca gcggctcagg agcccagggt ggcggaacca tctcctctgg 2040
 ctcttctgcc tccagtgtca cggtcactcg cagctaccgc agtggtgggg gcagtggggg 2100
 tggcagcttc ggggacaatc tggtcacccg ctctacctc ctgggcaact ccagccccc 2160
 aaccacagc cccagaact gcagcatcat gtaatctggg acctgccagg cagggtggg 2220
 ggtggaggct tcctgcgtcc tcctcacctc atgccaccc cctgccctgc acgtcatggg 2280
 agggggcttg aagccaaaga aaaataaccc ttggttttt ttcttctgta ttttttttc 2340
 taagagaagt tttttctac agtgggttta tactgaagga aaaacacaag caaaaaaaaa 2400
 aaaaaagcat ctatctcatc tatctcaatc ctaatttctc ctcccttctt tttccctgct 2460

33

tccaggaaac tccacatctg ccttaaaacc aaagagggct tcctctagaa gccaaaggaa 2520
 aggggtgctt ttatagaggc tagcttctgc ttttctgcc tgggctgctg cccccacccc 2580
 gggggaccct gtgacatggc gcctgagagg cagggcatag aggttctcc gccagcctcc 2640
 tctgggacgg caggcttcac tgccagggcc agcctccgag agggagagag agagagagag 2700
 gacagcttga gccgggcccc tgggtttggc ctgctgtgat tccactacac ctggctgagg 2760
 ttcctctgcc tgccccgcc ccagtcacca cccctgccc cagccccggg gtgagtccat 2820
 tctcccaggc accaagctgc gcttgctttt ctgtatttta tttagacaag agatgggaat 2880
 gagggtggag gtggaagaag ggagaagaaa ggtgagtttg agctgccttc cctagcttta 2940
 gaccctgggt gggctctgtg cagtcactgg aggttgaagc caagtggggt gctgggagga 3000
 gggagaggga ggtcactgga aaggggagag cctgctggca cccaccgtgg aggaggaagg 3060
 caagaggggg tggaggggtg tggcagtggt tttggcaaac gctaaagagc ccttgccctcc 3120
 ccatttccca tctgcacccc ttctctctc cccaaatcaa tacactagtt gtttctaaaa 3180
 aaa 3183

<210> 35
 <211> 207
 <212> DNA
 <213> Homo sapiens

<400> 35
 ccagggtgtt ggcgttttcc acagtaactg tgtatgttcc agcatctgtg tcactctcat 60
 cgttgatggc cagagccgc atcaagccaa tgacgcctgg cacaattcgg ccaggtttct 120
 ccaccacaat cttgccatcc ttctccaga ccacgtcagc ctctttgttt aactgcagc 180
 tcaagtacaa tggctgtcct ttgacca 207

<210> 36
 <211> 253
 <212> DNA
 <213> Homo sapiens

<400> 36
 atttattaca ttttttcatg cactgtcaag tttatcctcc gtccectaac ttctctacag 60
 gataccctt tctggtttgg ttcattgaaa tctgcaggga aagagctgcc ttcaaactcc 120
 tttgcttacc tcttccaaca ccttggaactc ttgaccgatt ttaccatctc aggtttcaga 180
 gccaggagag agccttgctc catcctgagc tgttcacccc catgggtatt ttctgccttt 240
 ctattccctc ttc 253

<210> 37
 <211> 687
 <212> DNA

34

<213> Homo sapiens

<400> 37

tgagccgccc ccgaggattc agcagcctcc cccttgagcc ccctcgcttc ccgacgttcc 60
gttccccctt gcccgcttc tcccgcacc gccgcgcgc ccttcgcag gccggtttcc 120
accgagggaa aggaatcgta tcgtatgtcc gctatccaga acctccactc tttcgacccc 180
tttgctgatg caagtaaggg tgatgacctg cttcctgctg gcactgagga ttatatccat 240
ataagaattc aacagagaaa cggcaggaag acccttacta ctgtccaagg gatcgctgat 300
gattacgata aaaagaaact agtgaaggcg ttttaagaaa agtttgctg caatgggtact 360
gtaattgagc atccggaata tggagaagta attcagctac agggtgacca acgcaagaac 420
atatgccagt tcctcgtaga gattggactg gctaaggacg atcagctgaa gggtcatggg 480
ttttaagtgc ttgtggctca ctgaagctta agtgaggatt tccttgcaat gagtagaatt 540
tcccttctct cccttgctac aggtttaaaa acctcacagc ttgtataatg taaccatttg 600
gggtccgctt ttaacttga ctagtgtaac tccttcagtc aataaactga aaagagccat 660
gctgtctagt cttgaagtcc ctcatctt 687

<210> 38

<211> 609

<212> DNA

<213> Homo sapiens

<400> 38

ggcgccgggg cccactgctc tgggctcccc caggagagga gcagagtctc gccagtgct 60
cctggagggg tgggagtggg gcctggcatt ctgaacacat ctctgagggg tgggattaat 120
aagacggctc ctgtgcctcc tgctccaga tcctgactgc tgtcatggcg tgccctctgg 180
agaaggccct ggatgtgatg gtgtccacct tccacaagta ctccggcaaa gagggtgaca 240
agttcaagct caacaagtca gaactaaagg agctgctgac ccgggagctg cccagcttct 300
tggggaaaag gacagatgaa gctgctttcc agaagctgat gagcaacttg gacagcaaca 360
gggacaacga ggtggacttc caagagtact gtgtcttcct gtcctgcac gccatgatgt 420
gtaacgaatt ctttgaaggc ttcccagata agcagcccag gaagaaatga aaactcctct 480
gatgtggttg gggggtctgc cagctggggc cctccctgtc gccagtgggc actttttttt 540
ttccaccctg gctccttcag acacgtgctt gatgctgagc aagttcaata aagattcttg 600
gaagtttta 609

<210> 39

<211> 2539

<212> DNA

<213> Homo sapiens

<400> 39

ccccctacat ggttctgctg gagagcaagc attttaccag ggatttaatg gagaagctga 60
aaggagagaac cagccgaatt gctggctctg cagtgtcctt gaccaagccc agtcctgcct 120
caggacatct ctccatagtg acagtgccca aatgatgggt ttgggtgtta ctccaattcc 180
tatgggccag agtttgctca ctgcagagaa atacagtggg attcgctggg caatgggttg 240
gcttatgaag acttttagttt ccccatcttt cttcttgaag atgaaaatga aaccaaagtc 300
atcaagcagt gctatcaaga tcacaacctg agtcagaatg gctcagcacc aaccttccca 360
ctatgtgccca tgcagctctt ttcacacatg catgctgtca tcagcactgc cactgcatg 420
cggcgagtc catccaaagc accttcagca tcaaccaga aatcgtctgt gacccctgt 480
ctgattacaa tgtgtggagc atgctaaagc ctataaatac aactgggaca ttaaagcctg 540
acgacaggggt tgtgtgtgct gccacccggc tggatagtcg ttccttttcc tggaaatgtgg 600
ccccaggggc tgaagcgca gtggcttcct ttgtcaccca gctggctgct gctgaagctt 660
tgcaaaagc acctgatgtg accaccctgc ccgcaatgt catgtttgtc ttctttcaag 720
gggaaacttt tgactacatt ggcagctcga ggatgggtcta cgatatggag aagggaagt 780
ttcccgctga gttagagaat gttgactcat ttgtggagct gggacaggtg gccttaagaa 840
cttcattaga gctttggatg cacacagatc ctgttttca gaaaaatgag tctgtacgga 900
accaggtgga ggatctcctg gccacattgg agaagagtgg tgctggtgct cctgctgtca 960
tcctcaggag gccaaatcag tcccagctc tcccaccatc tccctgcag cgatttcttc 1020
gagctcgaaa catctctggc gttgttctgg ctgaccactc tgggtgccttc cataacaaat 1080
attaccagag tatttacgac actgctgaga acattaatgt gagctatccc gaatggctga 1140
gccctgaaga ggacctgaac tttgtaacag aactgccaa ggccctggca gatgtggcca 1200
cgggtgctggg acgtgctctg tatgagcttg caggaggaac caacttcagc gacacagttc 1260
aggctgatcc ccaaacggtt acccgctgc tctatgggg tccctgattaa agccaacaac 1320
tcatgggtcc agtctatcct cagggcagga cctaaggtcc tacttgggtg acgggcctct 1380
tcaacattac atcgctgtct ccagcccccac caacaccact tatgtgtgac agtatgcctt 1440
ggcaaatttg actggcacag tggtaacct caccgagag cagtgccagg atccaagtaa 1500
agtcccaagt gaaaacaagg atctgtatga gtactcatgg gtccagggcc ctttgcattc 1560
taatgagacg gaccgactcc cccggtgtgt gcgttctact gcacgattag ccagggcctt 1620
gtgctcctgc ctttgaactg agtcagtggg gctctactga atactctaca tggactgaga 1680
gccgctggaa agatatccgt gcccgatat ttctcatcgc cagcaaagag cttgagttga 1740
tcaccctgac agtgggcttc ggcacctca tcttctccct catcgtcacc tactgcatca 1800
atgccaaagc tgatgtcctt ttcattgtc cccgggagcc aggagctgtg tcatactgag 1860

36

gaggacccca gcttttcttg ccagctcagc agttcacttc ctagagcatc tgtcccactg 1920
 ggacacaacc actaatttgt cactggaacc tccctgggccc tgtctcagat tgggattaac 1980
 ataaaagagt ggaactatcc aaaagagaca gggagaaata aataaattgc ctcccttcct 2040
 cegctccctt tcccacacac cccttcccca tttctcttc cttctctact catgccagat 2100
 ttgaggatta caaatagaag cttcttgctc ctgtttaaact ccctagttac ccaccctaact 2160
 ttgcccttca ggacccttct acttttttct tccctgccctg tacctctctc tgcctctcac 2220
 cccacccctt gtaccagacc accttcctga ctgggaagga cataaaaggt ttaatgtcag 2280
 ggtcaaaact cattgagccc ctgaggacag gggcatctct gggctgagcc tactgtctcc 2340
 tccccactgt cctttctcca ggccctcaga tggcacatta gggtagggct gctgcgggtg 2400
 ggtatccac ctccagccca cagtgtcag ttgtactttt tattaagctg taatatctat 2460
 ttttgtttt gtctttttcc ttattctctt ttgtaaatat atatataatg agtttcatta 2520
 aaatagatta tcccacacg 2539

<210> 40
 <211> 3146
 <212> DNA
 <213> Homo sapiens

<400> 40

ggagaaggag ctacctcccc acctggggga actgaccgtg gctgaggaga cctccagctc 60
 tctgcgcctg tcttgacgg tagccaggg cccctttgac tccctcgttg tccagtacag 120
 ggacacggac gggcagccca gggcagtgcc tgtggccgca gaccagcgca cagtaccgt 180
 agaggacctg gagcctggca agaaatacaa gttctctgctc tacgggctcc ttgggggaaa 240
 gcgcctgggc ccggtctctg ccttgggaat gacagcccca gaaggagaca caccagcccc 300
 agagttagcc ccagaggccc ctgagcctcc tgaagagccc cgcctaggag tgctgaccgt 360
 gaccgacaca accccagact ccatgcgccc ctctgtggagc gtggcccagg gccctttga 420
 ttccttcgtg gtccagtatg aggacacgaa cggcagccc caggccttgc tctgtgacgg 480
 cgaccagagc aagatcctca tctcaggccc gggcctcagc acccctaca ggttctctct 540
 ctatggcctc catgaaggga agcgccctggg cccctctca gctgagggca ccacagggct 600
 ggctcctgct ggtcagacct cagaggagtc aaqgcccgc ctgtcccagc tgtctgtgac 660
 tgacgtgacc accagttcac tgaggctcaa ctgggaggcc ccaccggggg ccttcgaact 720
 cttctctgct cgttttggg ttccatcac aagcactctg gagccgcatc cgcgtccact 780
 gctgcagcgc gagctgatgg tgccggggac gggcactctg gccgtgctcc gggacctgcg 840
 ttccgggact ctgtacagcc tgacactgta tgggctgcga ggacccaca aggccgacag 900
 catccaggga accgcccga cctcagccc agtctctggg agccccctg acctccaatt 960

cagtgaatac agggagacct cagccaaggt caactggatg cccccacat cccgggcgga 1020
cagcttcaaa gtctcctacc agctggcgga cggaggggag cctcagagtg tgcagggtga 1080
tggccaggcc cggaccaga aactccagg gctgatccca ggcgctcgct atgagggtgac 1140
cgtggtctcg gtccgaggct ttgaggagag tgagccctc acaggcttcc tcaccacggg 1200
tcctgacggg cccacacagt tgcgtgact gaacttgacc gagggattcg ccgtgctgca 1260
ctggaagccc cccagaatc ctgtggacac ctatgacgtc caggtcacag cccctggggc 1320
ccgcctctg caggcgga cccagcgag cgcggtggac taccctctgc atgaccttgt 1380
cctccacacc aactacaccg ccacagtgcg tggcctgcgg ggccccaacc tcacttcccc 1440
agccagcatc accttcacca cagggtctaga gggccctcgg gacttgagg ccaaggaagt 1500
gacccccgc accgcctgc tcacttgac tgagcccca gtccggcccg caggctacct 1560
gtcagcttc cacaccctg gtggacagaa ccaggagatc ctgctcccag gagggatcac 1620
atctcaccag ctcttggtc tcttccctc cacctctac aatggcacgg ctccaggcca 1680
tgtggggcca gagcctctg cgcctcgtt ccacctctt caccacgggt gggctgcgga 1740
tccccctccc cagggactgc ggggaggaga tgcagaacgg agccggtgcc tccaggacca 1800
gcaccatctt cctcaacggc aaccgcgagc gggccctgaa cgtgttttgc gacatggaga 1860
ctgatggggg cggctggctg gtgttccagc gccgcatgga tggacagaca gacttctgga 1920
gggactggga ggactatgcc catggttttg ggaacatctc tggagagtcc tggctgggca 1980
atgaggccct gcacagcctg acacaggcag gtgactactc catgcgcgtg gacctgcggg 2040
ctggggagca ggctgtgtc gccagtagc actccttcca cgtagactcg gctgcggagt 2100
actaccgct ccacttgag ggctaccacg gcaccgcagg ggactccatg agctaccaca 2160
gcggcagtgt ctctctgcc cgtgatcggg accccaacag cttgctcatc tcctgcgctg 2220
tctcctaccg aggggcctgg tggtagagga actgcccact acgccaacct caacgggctc 2280
tacgggagca cagtggacca tcaggagtg agctggtacc actggaaggg cttcaggttc 2340
tcggtgccct tcacggaaat gaagctgaga ccaagaaact ttcgctcccc agcgggggga 2400
ggctgagctg ctgccacct ctctcgacc ccagtatgac tgccgagcac tgaggggtcg 2460
ccccgagaga agagccaggg tccttcacca cccagccgct ggaggagacc ttctctgcca 2520
gcatctcgc agcactgtgt ttacagggg gaggggaggg gttcgtacgg gagcaataaa 2580
ggagaaactg aggtaccgg ctggcatcgg tctgccccca tcaactggtc tggcctgggc 2640
tgtggggccc catccccgg ggctgcagcc gcacttgaa aggctgcac ttgaggatga 2700
cactgcagtg gggcagggg tgcagggagg gcagggcgtc cccgagggc agcagcgtga 2760
aggcctgcag cagtcgggtc agcaccacga agagctccag gcgcgccagc ggctcggcca 2820
ggcacacgcg ggcaccgcag ccgaaggcca gagctctgga gttcttgct ggctccagga 2880

38

agcgatcagg ccagaactca tgtggcctct cccagaccgt ctcattccagg tgggcgcctt 2940
 ggagggttcgg aatgatgact gtgccctcag ggatgtcgta gccagagatg ctgctgggcc 3000
 gtgtggtgag gtggggcaag gctaagggca caacggggcg caggcgagc acctcggcga 3060
 tgggtggcatt gagcaagggc agccgtgcac ggtccttgta ggggacccgg gagctggagg 3120
 caccagggcc cagttcgtgg tctagc 3146

<210> 41
 <211> 2898
 <212> DNA
 <213> Homo sapiens

<223> unsure at all n locations
 <400> 41

acagagggac gtggtcactc tctgaaaagt tcaacttgag agacaaaatg cagtggacct 60
 ccttcctgct gctggcaggg ctcttctccc tctcccaggc ccagtatgaa gatgaccctc 120
 attggtggtt ccactacctc cgcagccagc agtccaccta ctacgatccc tatgaccctt 180
 acccgatga gacctacgag ccttaccctt atgggggtga tgaagggcca gcctacacct 240
 acggctctcc atcccccca gatcccccg actgccccca ggaatgcgac tgcccaccca 300
 acttccccac ggccatgtac tgtgacaatc gcaacctcaa gtacctgcc ttcgttcctt 360
 cccgcatgaa gtatgtgtac ttccagaaca accagatcac ctccatccag gaaggcgtct 420
 ttgacaatgc cacagggtg ctctggattg ctctccacgg caaccagatc accagtata 480
 aggtgggcag gaaggcttc tccaagctga ggcacctgga gaggtgtac ctggaccaca 540
 acaacctgac cggatgccc ggtcccctgc ctcgatccct gagagagctc catctcgacc 600
 acaaccagat ctacagggtc cccaacaatg ctctggaggg gctggagaac ctacggcct 660
 tgtacctcca acacaatgag atccaggaag tgggcagttc catgaggggc ctccggtcac 720
 tgatcttgct ggacctgagt tataaccacc ttcggaaggt gcctgatggg ctgccctcag 780
 ctcttgagca gctgtacatg gagcacaaca atgtctacac cgtccccgat agctacttcc 840
 ggggggcgcc caagctgctg tatgtgcggc tgtcccacaa cagtctaacc aacaatggcc 900
 tggcctccaa caccttcaat tccagcagcc tccttgagct agacctctcc tacaaccagc 960
 tgcagaagat cccccagtc aacaccaacc tggagaacct ctacctcaa ggcaatagga 1020
 tcaatgagtt ctccatcagc agcttctgca cgtgggtgga cgtcgtgaac ttctccaagc 1080
 tgcagggtgct gcgcctggac gggaacgaga tcaagcgag gncatgcct gccgacgcgc 1140
 ccctctgcct gcgccttgcc agcctcatcg agatctgagc agccctggca cgggtactg 1200
 ggcggagagc ccccgaggca tttggcttga tggtttggtt tggcttttgc tggaggtcc 1260
 aggatggacc atgtgacaga agtccacggg caccctctgt agtcttctt cctgtagggtg 1320
 gggtagggg gggcgatcag ggacaggcag ccttctgctg aggacatagg cagaagctca 1380

ctcttttcca gggacagaag tgggtggtaga tggaaggatc cctggatgtt ccaaccccat 1440
 aaatctcacg gctcttaagt tcttcccaat gatctgaggt catggaactt caaaagtggc 1500
 atgggcaata gtatataacc atacttttct aacaatccct ggctgtctgt gagcagcact 1560
 tgacagctct cctctgtgac tgggctgggc gtgcagttac tctgggctcc catttggtgc 1620
 ttctcaaat atacctctg cccagctgcc tctctgaaa tccacttcac ccactccact 1680
 ttctccaca gatgcctctt ctgtgcctta agcagagtca ggagacccca aggcattgta 1740
 gcatctgccc agcaacctgt ggagacaacc cacactgtgt ctgaggggtga aaggacacca 1800
 ggagtcactt ctatactcc ctaacctcac ccctggaaag ccaccagatt ggaggtcacc 1860
 agcatgatga taatattcat gacctgatgt gggaggagac agccaacctc aggccttagat 1920
 caatgtatag ggctatatct tggcagctgg gtagctcttt gaagtggtat aagacttcag 1980
 aagaggaaag gccagacttt gcttaccatc agcatctgca atgggcaaaa cacacctcaa 2040
 attggctgag ttgagaaagc agccccagta gttccattct tgcccagcac tttctgcatt 2100
 ccaaacagca tcctacctgg ggtttttatc cacaaggta gcggccacat ggtttttaaa 2160
 gtatgagaaa cacagtttgt cctctccttt tatccaagca ggaagattct atacctgat 2220
 ggtagagaca gactccaggg cagccctggg acttgctagc ccaaagaagg aggatgtggt 2280
 taatctgttt cacctggttt gtcctaaggc catagttaa aagtaaccagc tctggctggg 2340
 gtccgtgaag cccaggccag gcagccaaat ctgggcctgt gctgggcata caacctctg 2400
 ctttcacatc tctgagctat atcctcatta gtgaagggtg cttttgcttt atagtttggc 2460
 tggggagcac ttaattcttc ccatttcaaa aggtaatgtt gcctggggct taacctacct 2520
 gccctttggg caaggttggg acaagccat ctgggcagtc aggggcaagg actgttggag 2580
 gagagttagc ccaagtatag gctctgcca gatgcatca catccctgat actgtgtatg 2640
 ctttgaagca ctttccctga gaagggaaga ggggatcttt ggactaggtt cttggctcca 2700
 gacctggaat ccacaaaagc caaaccagct catttcaaca aaggagctcc gatgtgaggg 2760
 gcaaggctgc cccctgcccc agggctcttc agaaagcatc tgcattgtga caccatcatg 2820
 cctttataaa ggatccttat tacaggaaaa gcatgagtgg tggctaacct gaccaataaa 2880
 gttatcttat gattgcaa 2898

<210> 42
 <211> 854
 <212> DNA
 <213> Homo sapiens

 <223> unsure at all n locations
 <400> 42

ttcggcacag cgnnggggata caactctgga gtcctctgag agagccacca aggaggagca 60

40

```

ggggagcgcac ggccggggca gaagttgaga ccaccagca gaggagctag gccagtccat 120
ctgcatttgt caccaagaa ctcttaccat gaagaccctc ctactgttgg cagtgatcat 180
gatctttggc ctactgcagg cccatgggaa tttggtgaat tcccacagaa tgatcaagtt 240
gacgacagga aaggaagccg cactcagtta tggcttctac ggctgccact gtggcgtggg 300
tggcagagga tcccccaagg atgcaacgga tcgctgctgt gtcactcatg actgttgcta 360
caaacgtctg gagaaacgtg ggatgtgggc accaaatttc tgagctacaa gtttaggcaa 420
ctcggggagc agaatcacct gtgcaaaaca ggactcctgc agaagtcaac tgtgtgagtg 480
tgataaggct gctgccacct gttttgctag aaacaagacg acctacaata aaaagtacca 540
gtactattcc aataaacact gcagaggag caccctcgt tgcctgagtc cctcttcct 600
ggaaaacctc caccagtc tgaatttccc tctctcatc cctccctccc taccctaacc 660
aagttccttg gccatgcaga aagcatccct caccatcct agaggccagg caggagccct 720
tctataccca ccagaatga gacatccagc agatttccag ccttctactg ctctcctcca 780
cctcaactcc gtgcttaacc aaagaagctg tactccgggg ggtctcttct gaataaagca 840
attagcaaat catg 854

```

```

<210>      43
<211>      471
<212>      DNA
<213>      Homo sapiens

```

```

<400>      43

caataccatg aagaggaggc tcaggcagct cttaccacat gatacaagag ccggctggtg 60
gaagagtggtg gaccagaaag agaatttgct gaagaggaga agggaaaaaa aaacaccaa 120
aaaaaaaaata aaaaaatcca cacacacaaa aaaacctgcg cgtgagggggg gagggaaaagc 180
agggcctttt aaaaaggcaa tcacaacaac ttttgctgcc agggatgccc ttgctttggc 240
tgagaggatt tctgttgga agttgctgga ttatagttag gagttcccc accccaggat 300
ocgaggggca cagcgcggcc cccgactgtc cgtcctgtgc gctggccgcc ctcccaaagg 360
atgtacccaa ctctcagcca gagatggtgg aggccgtcaa gaagcacatt ttaaacatgc 420
tgcacttgaa gaagagaccg gatgtcacc agccgggtacc caaggcggcg c 471

```

```

<210>      44
<211>     1411
<212>      DNA
<213>      Homo sapiens

```

```

<400>      44

gccactgttc tgagaatttg tgagcagccc ctaacaggct gttacttcac tacaactgac 60
gatatgatca tcttaattta cttatttctc ttgctatggg aagacactca aggatgggga 120

```

41

ttcaaggatg gaatTTTTca taactccata tggcttgaac gagcagccgg tgtgtaccac 180
 agagaagcac ggtctggcaa atacaagctc acctacggca gaagctaagg cgggtgtgtga 240
 atttgaaggc ggccatctcg caacttaca gcagctagag gcagccagaa aaattggatt 300
 tcatgtctgt gctgctggat ggatggctaa gggcagagtt ggatacccca ttgtgaagcc 360
 agggcccaac tgtggatttg gaaaaactgg cattattgat tatggaatcc gtctcaatag 420
 gagtgaaga tgggatgcct attgctaca cccacacgca aaggagtgtg gtggcgtctt 480
 tacagatcca aagcaaattt ttaaatctcc aggcttccca aatgagtacg aagataacca 540
 aatctgctac tggcacatta gactcaagta tggtcagcgt attcacctga gttttttaga 600
 ttttgacctt gaagatgacc cagggtgctt ggctgattat gtgaaatat atgacagtta 660
 cgatgatgtc catggctttg tgggaagata ctgtggagat gagcttccag atgacatcat 720
 cagtacagga aatgtcatga ccttgaagtt tctaagtgtg gcttcagtga cagctggagg 780
 tttccaaatc aaatatgttg caatggatcc tgtatccaaa tccagtcaag gaaaaaatc 840
 aagtactact tctactggaa ataaaaactt tttagctgga agatttagcc acttataaaa 900
 aaaaaaaaa gatgatcaaa acacacagtg tttatgttgg aatcttttgg aactcctttg 960
 atctcactgt tattattaac atttatttat ttttttcta aatgtgaaag caatacataa 1020
 tttagggaaa attggaaaat ataggaaact ttaaacgaga aaatgaaacc tctcataatc 1080
 ccactgcata gaaataacaa gcgttaacat tttcatattt ttttctttca gtcatttttc 1140
 tttttgtggt atatgtatat atgtacctat atgtatttgc atttgaaatt ttggaatcct 1200
 gctctatgta cagttttgta ttatactttt taaatcttga actttataaa cattttctga 1260
 aatcattgat tattctacaa aaacatgatt ttaaacagct gtaaaatatt ctatgatatg 1320
 aatgttttat gcattattta agcctgtctc tattgttggga atttcaggtc attttcataa 1380
 atattgttgc aataaatatc cttgaacaca c 1411

<210> 45
 <211> 1877
 <212> DNA
 <213> Homo sapiens

<400> 45

gttcttgctt agtgagcaga tccagggggg tgtgatctcc gtgattaacc tggagcctag 60
 aactggcttc ttgtccaacc ctagggcctg gggccgcttt gacagtgtca tcacaggccc 120
 caacggggcc tgtgtggcct gccttctgtg atgaccagtc ccctgatgcc tactctgcct 180
 atgtcttggc aagcctggct ggggaggaac tgcaagcagt gggagtcttc tcctaaattc 240
 aacccaaatg caattggcgt ccctcagccc tatctcaaca agctcaacta ccgtcggacg 300
 gaccatgagg atccacgggt taaaaagaca gctttccaga ttagcatggc ccaagccaag 360

42

gcccactca gctgaggaga gcaatgggccc catctatgcc ttgagaacc tccgggcatg 420
 tgaagaggca ccacccagtg cagcccactt ccggttctac cagattgagg gggatcgata 480
 tgactacaac acagtcccc tcaacgaaga tgacctatg agctggactg aagactatct 540
 ggcattggtg ccaaagccga tggattcag ggcctgctat atcaaggga agattgtggg 600
 gccactggaa gtgaatgtgc gatcccgcaa catggggggc actcatcgcc ggacagtggg 660
 gaagctgtat ggaatccgag atgtgaggag cactcgggac agggaccagc ccaatgtctc 720
 agctgcctgt ctggagtca agtgagtggt gatgctctat gatcaggacc gtgtggaccg 780
 caccctggtg aagggtatcc ccagggcag ctgccgtcga gccagtgtga accccatgct 840
 gcatgagtac ctggtcaacc acttgccact tgcagtcaac aacgacacca gtgagtacac 900
 catgctggca cccttggacc cactgggcca caactatggc atctacactg tcaactgacca 960
 ggaccctcgc acggccaagg agatcgcggt tccgctggg ctttgatggc acatccgatg 1020
 gctcctccag aatcatgaag agcaatgtgg gatagccct caccctcaac tgtgtagaga 1080
 ggcaagtagg ccgccagagt gcctccagt accctcaaa caccacagcc cagtcccttg 1140
 ctgcaggcac tgtccaagga agagtgcctt caggaggca gcagcgagcg agcaggggtg 1200
 gccagcgcca gagtggagtg gtggcctctc tgagattctc tagagtgtct caacagcccc 1260
 tgatcaacta agttttgtgg tacttcaccc tctctgccc tcatttcagtg tgacagccat 1320
 tgtgagactg atgcacaaac tgtcacttgg ttaatttaag cactctgtt ttcgtgaatt 1380
 tgcttgtttg tttcttcag cctttactta ctttgcaca tgcactgat tggcacgtgg 1440
 cccccacaat ggcacaataa agcccccttg taaactgtt ctttaaatga aacacaagaa 1500
 attggccact ggtaaaactc tgcagcttca actgtacttc atttaatgcc attaatgcaa 1560
 atatacttcc tcttcttttt gcatggtttt gccacacct gcaatagtga taatctgatg 1620
 ctgaagatca aataaccaat ataaagcata tttcttggcc ttgctccaca ggacataggc 1680
 aaggccttga tcatagtcca tacatataaa ttttggcgaa ataaagaaat aaaacacaat 1740
 acttttactt gaaatgtaaa taacttattt tttcttctgc taaatttggg attctagtgc 1800
 acattcaaag ttaagctatt aaatataggg tttatcatagt tctctacca agtctggaaa 1860
 agaacatctc ctggtat 1877

<210> 46
 <211> 167
 <212> DNA
 <213> Homo sapiens

<400> 46

atcaaaaaa tcaactccctc tccctcccta acagtgaaaa gagagaaggg agactctatt 60
 taagattccc aaacctaatt atcatctgaa tcccgggcta agaattgcaga cttttcagac 120

43

tgaccccgaga aattctggcc cagccaatct agaggcaagc ctggcca 167

<210> 47
<211> 1689
<212> DNA
<213> Homo sapiens

<400> 47

cccgctccg ccacctttct tgggtggctc tccgcctcgt cctccctccg agggccggtg 60
gtacattcct agtgactcca agcgcttaaa aggggcccgg gaggatgaac cccacagatc 120
tgaacctgat ttgtgtgtgc accgcgtctc cagcgatccc ggatccactg cgctgccagg 180
gcgctgggg gtgggtctct tgcgtctctc gcgacgacat ccttacgttt cggcactcta 240
atgctgggtt tgtgcgtgtg tgtctgctta gcggtctagc gggctgttag gctccctcgc 300
ccccagctcc ttggctcgtc cagctcctcc accgcagccc agcagtgaga cgcgcgcgca 360
gccagctccc cagagatgg aacagaccga agtgctgaag ccacggaccc tggctgatct 420
gatccgcatc ctgcaccagc tctttgccgg cgatgaggtc aatgtagagg aggtgcaggc 480
catcatggaa gcctacgaga gcgacccac cgagtgggca atgtacgcca agttcgacca 540
gtacaggtat acccgaaatc ttgtggatca aggaaatgga aaatttaatc tgatgattct 600
ctgttggggg gaaggacatg gcagcagtat tcatgatcat accaactccc actgctttct 660
gaagatgcta cagggaatc taaaggagac attatttgcc tggcctgaca aaaaatccaa 720
tgagatggtc aagaagtctg aaagagtctt gagggaaaac cagtgtgcct acatcaatga 780
ttccattggc ttacatcgag tagagaacat cagccatagc gaacctgctg tgagccttca 840
cttgtagagt ccaccttttg atacatgcc a tgcctttgat caaagaacag gacataaaaa 900
caaagtcaca atgacattcc atagtaaatt tggaaacaga actccaaatg caacttcggg 960
ctcgtgggag aacaactaag gggcaccaaa ccctctgagg ttttacttta aggttcgctg 1020
tatgtttgcc ttggacaaaa aggctaccta ccacgtgcta tccagtaata tacttaaata 1080
agccaatact tagatctact gtaaggcaga tgtaattat aaggcatata gtaagcaaat 1140
agtgcctca gctactgcag aagaaaagtc cactgagga aaagaaagtc ttgtgatttt 1200
taaaggcaag ttttcaagtg ctctcatagt tctatcctct aattccatta aatccatact 1260
aggagcgtca gtgagggttt tcatagcttt tggaaatact ttggtctctg aactgtaatt 1320
agcaagaagt aaaaacagaa acgtcaaacg tcaaatgttt gctttgttac ctggaggact 1380
aatgtagat gtcttttagta tactttgtat gttcttaata ttggaagata attttgtgaa 1440
tctgtagatt ttattttttc agtcttacct taaaaatttc ttttctatga ataatagagg 1500
aacttacggc actctgccat ttgttaatga aaggaagtgc agaggattta gaaaagtaca 1560
tgatccccag accacaacaa accaaaacat aaactcatgt ctgtgtccca tggctcatgt 1620

44

caaaagatttt gtactgctaa aattaccaaa taattttaa aaagtggatt tgaacacaaa 1680

aaaaaaaaa 1689

<210> 48
<211> 184
<212> DNA
<213> Homo sapiens

<400> 48

agaaaacaat gaagaatcga atgaagatga agactctgag gctgagaata ccacactttc 60

tgctacaaca ctgggctatg gagaggacgc cagcctggc acaggggtata caggggttagc 120

tgcaatccag ctcccaaga aggctgggga tataacaaac aaagctacaa aagagaagga 180

aagt 184

<210> 49
<211> 259
<212> DNA
<213> Homo sapiens

<400> 49

cctggccccc tgggtcctcc tggcctgacg ggtcctgcag gtgaacctgg acgagagggg 60

agccccgggtg ctgatggccc ccctggcaga gatggcgctg ctggagtcaa gggatgatcgt 120

ggtagactg gtgctgtggg agctcctgga gcccctgggc cccctggctc ccctggcccc 180

gctgggtcaa ctggcaagca aggagacaga ggagaagctg gtgcacaagg ccccatggga 240

ccctcaggac cagctggag 259

<210> 50
<211> 245
<212> DNA
<213> Homo sapiens

<400> 50

gagagaaggg ccaccaggt ctcatggac tgattgggcc cccgggtgag caggagagaga 60

agggagatcg gggacttctt gggcctcagg gctcccttgg gcagaagggt gagatgggta 120

tcccaggagc atccggcccc attggctctg gaggtcccc cggcctcccc ggacctgctg 180

gccccaaagg agccaaagga gccacaggcc caggcggacc caaggagag aagggtgtgc 240

agggc 245

<210> 51
<211> 515
<212> DNA
<213> Homo sapiens

<400> 51

45

cttgcagaga aagagtcttt tgtgcagcac ccttttaaagg gtgactcgte ccacttgtgt 60
 tctctctcct ggtgcagagt tgcaagcaag ttatcagag tatcgccatg aagttcgtec 120
 cctgccttct gctggtgacc ttgtcctgcc tggggacttt gggtcaggcc ccgaggcaaa 180
 agcaaggaag cactggggag gaattccatt tccagactgg agggagagat tcctgcacta 240
 tgcgtcccgag cagcttgggg caaggtgctg gagaagtctg gcttcgcgtc gactgccgca 300
 acacagacca gacctactgg tgtgagtaca gggggcagcc cagcatgtgc caggctttcg 360
 ctgctgaccc caaatcttac tggaatcaag ccctgcagga gctgaggcgc cttcaccatg 420
 cgtgccaggg ggccccggtg cttaggccat ccgtgtgcag ggaggctgga ccccaggccc 480
 atatgcagca ggtgacttcc agcctcaagg gcagc 515

<210> 52
 <211> 281
 <212> DNA
 <213> Homo sapiens

<400> 52
 gcccggggcc ctggacgatg tggagaacct cgccaaattc cacgtggaca ggaaccagct 60
 gtccagctac ccctcagctg ccctgagcaa gctacgggtg gtggaggagc tgaagctgtc 120
 ccacaacccc ctgaaaagca tcccggacaa tgccttccag tcctttggca gatacctgga 180
 gaccctctgg ctggacaaca ccaacctgga gaagttctca gatggtgcct tcctgggtgt 240
 aaccacgctg aaacacgtcc atttgagaa caaccgcttg a 281

<210> 53
 <211> 252
 <212> DNA
 <213> Homo sapiens

<400> 53
 gggacagatc ccagggtgcc caggagctct ccaagtgcct cactcctccc gccgcaaaca 60
 tgacagagaa ctccgacaaa gtcccatctg ccctgggtggg acctgatgac gtggaattct 120
 gcagcccccc ggcgtacgct acgtgacgg tgaagccctc cagccccgcg cggctgctca 180
 aggtggggag cgtggctctc atttcgggag ctgtgctgct gctctttggg gccatcgggg 240
 ccttctactt aa 252

<210> 54
 <211> 2723
 <212> DNA
 <213> Homo sapiens

<400> 54
 gacatagctt ttctcattca ccctcccact tggggctaata gcacagacat gaacatctat 60

tgaggaaaac cacaaaaaac ttcaaaacag ctacaacggg aaaaagagag ttttgtccca 120
cagtcagcag gccactagtt tattaacttc cagtcacctt gatTTTTgct aaaatgaaga 180
ctctgcagtc tacacttttc ctgttactgc ttgtgcctct gataaagccc aggaccacc 240
aaccagcag gactcacgca ttatctatga ttatggaaca gataatTTTg aagaatccat 300
atTTtagccaa gattatgagg ataaatacct ggatggaaaa aatattaagg aaaaagaaac 360
tgtgataata cccaatgaga aaagtcttca attacaaaa gatgaggcaa taacaccatt 420
acctcccaag aaagaaatg atgaaatgcc cacgtgtctg ctgtgtgttt gTTtaagtgg 480
ctctgtatac tgtgaagaag ttgacattga tgctgtacca cccttaccaa aggaatcagc 540
ctatctttac gcacgattca acaaaattaa aaagctgact gccaaagatt ttgcagacat 600
acctaactta agaagactcg atTTTtacagg aaatttgata gaagatatag aagatggtac 660
TTTTTcaaaa ctttctctgt tagaagaact ttcacttgct gaaaatcaac tactaaaaact 720
tccagttctt cctcccaagc tcactttatt taatgcaaaa tacaacaaaa tcaagagtag 780
gggaatcaaa gcaaatgcat tcaaaaaact gaataacctc accttctct acttggaacca 840
taatgccctg gaatccgtgc ctcttaattt accagaaagt ctacgtgtaa ttcattctca 900
gttcaacaac atagcttcaa ttacagatga cacattctgc aaggctaag acaccagtta 960
catccgggac cgcattgaag agatacgctt ggagggaat ccaatcgtcc tgggaaagca 1020
tccaaacagt tttatttgct taaaaagatt accgataggg tcatactttt aacctctatt 1080
ggtacaacat ataaatgaaa gtacacctac actaatagtc tgtctcaaca atgagtaaag 1140
gaacttaagt attggTTTaa tattaacctt gtatctcatt ttgaaggaa ttaatatTTT 1200
aagcaaggat gttcaaaatc ttacatataa taagtaaaa gtaagactga atgtctacgt 1260
tcgaaacaaa gtaatatgaa aatattTaaa cagcattaca aaatcctagt ttatactaga 1320
ctaccattta aaaatcatgt ttttatataa atgcccaat ttgagatgca ttattcctat 1380
tactaatgat gtaagtacga ggataaatcc aagaaacttt caactctttg cctttcctgg 1440
cctttactgg atcccaaaag catttaaggt acatgttcca aaaactTTTg aaagctaaat 1500
gtttcccatg atcgctcatt ctctttttat gattcatacg ttattcctta taaagtaaga 1560
actttgTTTT cctcctatca aggagctat tttattaaat ttttactta gtctgagaaa 1620
tagcagatag tctcatattt aggaaaactt tccaaataaa ataaatgtta ttctctgata 1680
aagagctaag acagaaatgt tcaagttatt ttactttctg gtaatgtctt cagtaaaata 1740
ttttctttat ctaaataatta acatttctaag tctaccaaaa aaagttTtaa actcaagcag 1800
gccaaaacca atatgcttat aagaaataat gaaaagttca tccatttctg ataaagttct 1860
ctatggcaaa gtctttTcaa tacgagataa ctgcaaaata ttttcttttt atactacaga 1920
aatgagaatc tcatcaataa attagttcaa gcataagatg aaacagaat attctgtggT 1980

47

gccagtgcac actaccttcc caccataca catccatgtt cactgtaaca aactgaatat 2040
 tcacaataaa gcttctgagt aacactttct gattactcat gataaactga catggctaac 2100
 tgcaagaatt aaatcttcta tctgagagta ataatttatg atgactcagt ggtgccagag 2160
 taaagtttct aaaataacat tcctctcact tgtacccac taaaagtatt agtctacaca 2220
 ttacattgaa gttaaacaca aaattatcag tgttttagaa acatgagtcc ggactgtgta 2280
 agtaaaagta caaacattat ttccaccata aagtatgtat cgaaatcaag ttgtctctgt 2340
 gtacagaata catacttatt cccattttta agcatttgc tctgttttcc ctacctagaa 2400
 tgtcagatgt ttttcagtta tctcccccatt tgtcaaagtt gacctcaaga taacattttt 2460
 cattaagca tctgagatct aagaacacaa ttattattct aacaatgatt attagctcat 2520
 tcacttattt tgataactaa tgatcacagc tattatacta ctttctcgtt attttgtgtg 2580
 catgcctcat ttccctgact taaacctcac tgagagcgca aaatgcagct ttatactttt 2640
 tactttcaat tgcctagcac aatagttagt acatttgaat tgaatatata ataaatattg 2700
 caaaataaaa tccatctaaa tag 2723

<210> 55
 <211> 310
 <212> DNA
 <213> Homo sapiens

<400> 55
 ggcgcgcgcc gccgctgctg cccccagccc cggccccagg cgtcccagcc atggtcgcgc 60
 caatgctctt gctcagcctc ggccctcctg ctggctcgtc gccggcgctg gccgcctgcc 120
 ccagaaactg cactgcccac agcgacctgc agcacgtcat ctgcgacaag gtggggctgc 180
 agaagatccc caaggtgtca gagaagacca agtgcgtcaa cctacagcgc aacaacttcc 240
 cgggtgctggc tgccaattcg ttccgggcca tgccgaacct cgtgtcattg cacctgcagc 300
 actgccagat 310

<210> 56
 <211> 274
 <212> DNA
 <213> Homo sapiens

<400> 56
 atttatgaaa tcataaaacc tgcaacagcc aactcgaaat tccccgtgac cagtcttttg 60
 gacaccaggg acagcaatga gcctgactct cctgcacctc ctttgtctga ggcatagacc 120
 actgactgct tatggaaaag aacagataat gatatccgtc tctgcttcc acccaccact 180
 caatgtaact ttctgccatg aacataacca gccacacata aactgtctgc agaaaaggaa 240
 gttccatcct ataagcttgg caggaggata aaga 274

48

<210> 57
 <211> 153
 <212> DNA
 <213> Homo sapiens

<400> 57
 aattttaaga ttttaactta cacaaaaagt ccacttacaa gcatttatct catttacatg 60
 tattcacctt ttccatttct taatagtta tctagattac ttctgaaaac tgagatatta 120
 cacaaaacta atcattatct aaagttatct ccg 153

<210> 58
 <211> 225
 <212> DNA
 <213> Homo sapiens

<400> 58
 tgatggtaag ttgtttcagg cataaaattt gaaataaatt atgaggctcc atgatatgct 60
 atattggttt taccttcaga agaattttta gtttactca ggtttttcaa agctacgctg 120
 tcccccaaaa aacgaaacaa aacaaaaaaa caacctttt aagagttgat ggctactcat 180
 ttgatctgcc tcctctgctg aatcaattag gaattttttt ttttt 225

<210> 59
 <211> 448
 <212> DNA
 <213> Homo sapiens

<400> 59
 ggaagcgtcc aaagagggac ggctgtcagc cctggcttga ctgagaaccc accagctcat 60
 cccagacacc tcatagcaac ctattttatc aaagggggaa agaaacacct gagcagaatg 120
 gaatcattat ttttttccca aggagaaaac cggggtaaaag ggaggggaagc aattcaattt 180
 gaagtccttg tgaatgggct ttcagaaggc aattaaagaa atccactcag agaggacttg 240
 gggtgaaaact tgggtcctgt gggtttctga ttgtaagtgg aagcaggtct tgcacacgct 300
 gttggcaaat gtcaggacca gggttaagtga ctggcagaaa aacttccagg tggacaagc 360
 aacctcaggt ctgctgcaag cttggaagga tcttgagcgg ggagaaagct aacttgaaca 420
 tgacctgttg catttgcaa gttctagc 448

<210> 60
 <211> 59
 <212> DNA
 <213> Homo sapiens

<400> 60
 atgacattgg ttgcctcagc cctgaaaagc tatgtctctg cattcttagt tttctttgt 59

<210> 61
<211> 321
<212> DNA
<213> Homo sapiens

<223> unsure at all n locations
<400> 61

attaattgcc agtagttgta aggaggagtc agcatctagt gttactccct nnnnnnnnnn 60
nnnnnnnnnn nnnntccagg tactggctaa tggagctact gccacctcta aacccctcca 120
gccactaggc tgtgtcccac agtcagtgtc acccagtga caggcattac cccacatct 180
ggaaccagcc tggccccaag ggctacggca taactcagta ccaggtagag ttggcccccac 240
agagtacctt tccccagata tgcaacgcca gcgaaagacc aagcgcaaaa ccaaagagca 300
gctggctatc cttaaatcct t 321

<210> 62
<211> 252
<212> DNA
<213> Homo sapiens

<400> 62

tttcccta atttaaatta ttctttataa accagtagaa aagctttaac aacataacag 60
aaaaatggga aaagactatg aatagacggg acccagaaaa gcacatacaa ataagtggct 120
attttactac acctttactt tggaaaactt caaacctgta ctaaaataga atagggcagt 180
gaacctccct gcctgcaccc atcactcagc gtcaacattg atcaactcat gggcaatctt 240
gttttatcta tt 252

<210> 63
<211> 218
<212> DNA
<213> Homo sapiens

<400> 63

cacaagttaa aacttcccat gtataaaaac acttacattt taaaacatca ctgccaaactg 60
tgtgctcatg tgggagtaca gatgtgtata tacagacatg tacattttta aagacttggg 120
tgtctctgca gtgaagacaa tatgttttat ttttattcc atatacttct ctgtattttc 180
tatatttgct tcaataagct ggtgtaactt ttaatttt 218

<210> 64
<211> 235
<212> DNA
<213> Homo sapiens

<400> 64

gatcaaatcg gaaaggtaaa gatgaaatgc ttttcctgtt tcttgatttt tatctaccag 60

caataatatg aggcacactc gtaaagtaaa ggtttgcat atattttacaa ttaaactcta 120
 gaaaagcata attctgagct aaatattctg cctaaagaat ctctttcaca taatccttcc 180
 tggtcacttg ctccttgac tcacaatttg tttcttaatt cctatgcttt ttatc 235

<210> 65
 <211> 239
 <212> DNA
 <213> Homo sapiens

<400> 65

tgccgctttg ttgagccctt aaaataccac ctcctcatgt gtaaattgac acaatcacta 60
 atctggtaat ttaaacaatt gagatagcaa aagtgtttaa cagactagga taattttttt 120
 ttcataattg ccaaaatttt tgtaaaccct gtcttgtaa ataagtgtat aatattgtat 180
 tattaattta tttttacttt ctataccatt tcaaaacaca ttacctaag ggggaacca 239

<210> 66
 <211> 243
 <212> DNA
 <213> Homo sapiens

<400> 66

ggaaactcca ggctcctgg tttccctgg gcggggaaag agaagactga aacatctgtg 60
 tgacattcag atttttcaga ggtctgcca agggctctgg ttttattttg cttgaatata 120
 agttctgaca ggaaaggcca ccaggttgcg gggtcattga aaacaaagt gacagtttag 180
 attagcaggc actcaccatg gtccctcccc ctccctcagc atgaaaacca gcaggagaaa 240
 ttc 243

<210> 67
 <211> 250
 <212> DNA
 <213> Homo sapiens

<400> 67

gtctgtgtac catcttacct ggaatagaga ttgtgttaaa ttaacagatc atctgactga 60
 gaggtttttt tccccaaaa cagaagcaaa taaacattat tttgttcctt tggataact 120
 ttcattgaac agttatatag tgctttggaa gtatcaagtc ctgtgctaaa taaatgctgg 180
 agatacaaaa gccctgacc tcagaatgtc atagtcttg ggtaagaaaa aattcattct 240
 gtgcccagg 250

<210> 68
 <211> 213
 <212> DNA
 <213> Homo sapiens

<400> 68

cagggtgtgaa ccactgcacc tggcccaaaa tctcttgatt gatacagtc tctttatttt 60
 tcaagatcaa gttatgatac ctttaccac agtcatacat tcttttgaa ctttgcacaa 120
 tagtcatatg ttcttttaga actttacact tctattcttt attgccctgt attataattg 180
 cttgtatgcc tgactcctct acatgactgt atg 213

<210> 69

<211> 198

<212> DNA

<213> Homo sapiens

<400> 69

cataaaccta ctttatcatc ctctcctaaa gggaaaagag aagatttagc tagaataatt 60
 attaacagaa gatgtggaga tacagaagaa actagaaaat atctcacaat caatacatct 120
 ttcaagcagt caatcatttg tcaactcatat tgctttttta aaccagctt tacatggaag 180
 gaataaatgg aactccag 198

<210> 70

<211> 393

<212> DNA

<213> Homo sapiens

<400> 70

aaaaaaagga aaaaaaaaaat tgctttaagt catatagatt gtaccagcag ctctcacagt 60
 gtggactttg gacttctagg agtccccagg aaccttttag gggatgccta cgaggagggtc 120
 caaactgttt tcataagaac gctaagggtgc tatgtgcctt ttttaactcat tctctcacga 180
 gtgttcagtg gagttttcca gaggtctgt gacatgggtga catcactctg ataattagta 240
 gaatgtgtgt gtgtgtactt ttgttttcta gaatattgta aattgataga tttagggtat 300
 aaatatatgt gttttcagag attaactcag tttgtgccca gtgcttctac tgtgctctta 360
 ctggctattt tcatttatac ctgctgctga gtc 393

<210> 71

<211> 216

<212> DNA

<213> Homo sapiens

<400> 71

ctctacttgt atgaccctag gaatagattg gaatactgca gaggacacaa gctgaggcat 60
 gctaaacagc tgcttgagg tggaagcaag ttcagtcacc tactcagctt cctctctcca 120
 ccaccagtt cctccctcag tatcacatta tttttttctt ctgcttttca ttaacctaac 180
 tcattctatc agtacaacca ttttcttatt ctctaa 216

<210> 72
 <211> 166
 <212> DNA
 <213> Homo sapiens

<400> 72

caaataattta acagaactaa tggaactatt ttagtatgct ttccctggg ctggagtgtg 60
 ggctaagact ttatttaaata acaggatgga tgggtgtttg actgaagatg cctccaactt 120
 ttgctcttct gttttttatt tgatgtgctc aagcttctaa ttccct 166

<210> 73
 <211> 240
 <212> DNA
 <213> Homo sapiens

<400> 73

tgataggcag ctaaaactgt tatgccact gtgctcaatt tgaagcagaa ttcagtgaag 60
 aattattttt ccacattgaa acactttgca gacacaaata tctatgaaaa gatgctttgt 120
 cagccactgt gcctttttt ctgtgaagac tcaacggatg tgtgtgtttg tatgtttgtt 180
 aacagttaca tatgtttgta tgagtgtata tatatatctg tgtgtgtgta tctctaactg 240

<210> 74
 <211> 291
 <212> DNA
 <213> Homo sapiens

<400> 74

tggaccccc gctgaggagt cctgctcaag acacgggtcac tggatctgag aaacttccca 60
 ggggaccgca ttccagagtc agtgactctg tgaagcacc acatctacct cttgccacgt 120
 tcccacgggc ttgggggaaa gatggtgggg accaaggcct ggggtgttctc cttcctggtc 180
 ctggaagtca catctgtgtt ggggagacag acgatgctca cccagtcagt aagaagagtc 240
 cagcctggga agaagaacct cagcatcttt gccaagcctg ccgacaccct g 291

<210> 75
 <211> 283
 <212> DNA
 <213> Homo sapiens

<400> 75

ctccgccagc ctccgggaga ggagccgcac ccggccggcc cggccccagc cccatggacc 60
 tccgagcagg ggactgcgtg ggggatgtta gcgtgcctgt gcacgggtgct ctggcacctc 120
 cctgcagtgc cagctctcaa tcgcacaggg gaccagggc ctggccccctc catccagaaa 180
 acctatgacc tcaccegeta cctggagcac caactccgca gcttggctgg gacctatctg 240

aactacctgg gcccccttt caacgagcca gacccaacc etc 283

<210> 76
 <211> 139
 <212> DNA
 <213> Homo sapiens

<400> 76

ccttcgtgaa gtcgccaac ctctctgagc cccagtcatt gctagtaaga cctgcctttg 60

agttggtatg atgttcaagt tagataacaa aatgtttata cccattagaa cagagaataa 120

atagaactac atttcttgc 139

<210> 77
 <211> 669
 <212> DNA
 <213> Homo sapiens

<400> 77

ctggctggag cagcagctct gtcgatccca ggccagagac aaggcagaca aaggttcatt 60

tgtaaagaag ctccctccag caccctctct ctctctcttt tgcccaaaact caccagtgga 120

gtgtgagcat ttaagaagca tctctgcca agacaaaag gaaagaagaa aaagggccaa 180

aagccaaat gaaactgatg gtacttggtt tcaccattgg ggctaacttt gctgctagga 240

gttcaagcca tgcctgcaaa tcgctctctt tgctacagaa agatactaaa agatcacaac 300

tgtcacaacc ttccggaagg agtagctgac ctgacacaga ttgatgtcaa tgtccaggat 360

catttctggg atgggaaggg atgtgagatg atctgttact gcaacttcag cgaattgctc 420

tgctgcccc aagacgtttt ctttgacca aagatctctt tcgtgattcc ttgcaacaat 480

caatgagaat cttcatgtat tctggagaac accattcctg atttcccaca aactgcacta 540

catcagtata actgcatttc tagtttctat atagtgaat agagcataga ttctataaat 600

tcttacttgt ctaagacaag taaatctgtg ttaaacaagt agtaataaaa gtttaattcaa 660

tctaaaaaa 669

<210> 78
 <211> 486
 <212> DNA
 <213> Homo sapiens

<400> 78

ggacgccatc tctgaggccc aaggccacag tgaaatcaca gaagcaacac agctgggaaa 60

ggactcgtat gaagagctgg gaaaagccaa acccaccacc cgaccacag ccaaactac 120

ccagcctgga cccagggccc gaggggaatga ggaagcaaag aagaaggcct gggaacattg 180

ttggaaaccc ttccaggccc tgtgcgcctt tctcatcagc ttcttccgag ggtgacaggt 240

54

gaaagacccc tacagatctg acctctccct gacagacaac catctctttt tatattatgc 300
 cgcttttcaat ccaacgttct cacactggaa gaagagagtt tctaatacaga tgcaacggcc 360
 caaattcttg atctgcagct tctctgaagt ttggaaaaga aaccttcctt tctggagttt 420
 gcagagttca gcaatatgat aggggaacagg tgctgatggg cccaagagtg acaagcatac 480
 acaact 486

<210> 79
 <211> 752
 <212> DNA
 <213> Homo sapiens
 <223> unsure at all n locations
 <400> 79

ggggctacga gcccacagag gatggcacag ccttctggg gattctcggc cagtcaccgg 60
 gccccgcgc caccaccccc ancnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
 nnnnngccac tgctgcaccg gtctctgtag atggagatct caatctgggg tcggtggtta 180
 aggagagctg caagcccagc tgctgagcag ggggtgggac atgaaccagc ggatggagtc 240
 cagcagggga gtgggaaagt gggcttgtgc tgctgcctag atagtaggga tgtaaaggcc 300
 tgggagctag accctcccca agcccatcca tgacattac ttagctaaca attagggaga 360
 ctctgaaggc caggccctgt gctgggcaca tagctgtgat ctcagcagac agggtcgctg 420
 cctgatggc gtttacattc cagtgggtct atgaccata ttttaggaca cagatgtgcc 480
 caggaggggtg gtgtcactgc acaggaagta tgaggacttt agtgtcctga gttcaaatcc 540
 tgattcagga actcacaag ctatgtgacc ttacaccagt cacttaactt gttagccatc 600
 cattatcgca tctgcaaat ggggattaag aatagaatct tggggttagt gtggagatta 660
 gattaaatgt atgtaagaca cttggcaca aacctggnac atagtaaagg ctcaataaaa 720
 acaagtgcct ctcactgggc tttgtcaaca ca 752

<210> 80
 <211> 552
 <212> DNA
 <213> Homo sapiens
 <223> unsure at all n locations
 <400> 80

aaatatattc tcaacatttt cagtggagaa tttctgtaat ggcacctcaa atnttatact 60
 cttaaaaaan aacaataatt tgtgaattac caccaaaagg caatggcagt cctacattta 120
 agaatagagc tatgcaaaact ctgttaaaaa ctatgaggaa aacttatatt agaacttttg 180
 atatatata aaatactgat tatcttaatc acattttccc cagagataaa cattgagaga 240
 acgaaagcca aagtgtcatt taagagagat atatatgaaa aagtaacatt aatatataga 300

55

actttacccat caccagccgt agttgataga aaatattagt ttcagaatta ccctccttta 360
aaaaataaga gactatttgt tttcttttaa tttctatgaa taaaagaaat ttttaaaaac 420
tttaaaatTT taaatattag tcaaaatact ttttaagtcc tgagtgccta caggtagttg 480
ttaaaaaaat ttaaggcca ggcattggtg ctcgctcaca cctataatcc taggatctgg 540
gaggtcgagg ca 552

<210> 81
<211> 135
<212> DNA
<213> Homo sapiens

<400> 81
ttcactcttc aaatgtttgc ttctgttcc tgctaccctg aaccctgctg ttgaggggtt 60
ctagtgtcta caaggggaac gctgccacca cgaggaataa cacagtgtc ttacagcctg 120
ttccaagtgt ggctt 135

<210> 82
<211> 225
<212> DNA
<213> Homo sapiens

<400> 82
ggagaatgtg acatagattt gctggcacat gggtttccta tgagcaaacc ccagaattgg 60
acacacgtat ctggtgctgc attggaatca tccgaaaaaa ccaaggcttg cattgcata 120
ctatctgctg tctgctgaag gagccctgtc tgtgtgcccc aggaagtga atccttgcca 180
agggctgtcc ctgttgagg agatgaagga gccctgtcta tgtgc 225